
Article:
Adam, T., Griffiths, C.A. orcid.org/0000-0001-7203-0426, Leos-Barajas, V. et al. (5 more authors) (2019) Joint modelling of multi-scale animal movement data using hierarchical hidden Markov models. Methods in Ecology and Evolution, 10 (9). pp. 1536-1550. ISSN 2041-210X

https://doi.org/10.1111/2041-210x.13241

Methods in Ecology and Evolution

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

Timo Adam[1]*, Christopher A. Griffiths[2,5], Vianey Leos-Barajas[1,3], Emily N. Meese[4], Christopher G. Lowe[4], Paul G. Blackwell[2], David Righton[5], and Roland Langrock[1]

[1]Bielefeld University, Bielefeld, Germany

[2]University of Sheffield, Sheffield, UK

[3]Iowa State University, Ames, USA

[4]California State University, Long Beach, USA

[5]CEFAS, UK

June 12, 2019

Abstract

1. Hidden Markov models are prevalent in animal movement modelling, where they are widely used to infer behavioural modes and their drivers from various types of telemetry data. To allow for meaningful inference, observations need to be equally spaced in time, or otherwise regularly sampled, where the corresponding temporal resolution strongly affects what kind of behaviours can be inferred from the data.

2. Recent advances in biologging technology have led to a variety of novel telemetry sensors which often collect data from the same individual simultaneously at different time scales, e.g. step lengths obtained from GPS tags every hour, dive depths

*Corresponding author; email: timo.adam@uni-bielefeld.de.

obtained from time-depth recorders once per dive, or accelerations obtained from accelerometers several times per second. However, to date, statistical machinery to address the corresponding complex multi-stream and multi-scale data is lacking.

3. We propose hierarchical hidden Markov models as a versatile statistical framework that naturally accounts for differing temporal resolutions across multiple variables. In these models, the observations are regarded as stemming from multiple, connected behavioural processes, each of which operates at the time scale at which the corresponding variables were observed.

4. By jointly modelling multiple data streams, collected at different temporal resolutions, corresponding models can be used to infer behavioural modes at multiple time scales, and in particular help to draw a much more comprehensive picture of an animal's movement patterns, e.g. with regard to long-term vs. short-term movement strategies.

5. The suggested approach is illustrated in two real-data applications, where we jointly model i) *coarse*-scale horizontal and *fine*-scale vertical Atlantic cod (*Gadus morhua*) movements throughout the English Channel, and ii) *coarse*-scale horizontal movements and corresponding *fine*-scale accelerations of a horn shark (*Heterodontus francisci*) tagged off the Californian coast.

Keywords: animal movement modelling; biologging; state-space models; statistical ecology; temporal resolution; time series modelling.

1 Introduction

Movement ecology seeks to identify and understand the key patterns in animals' movements through space, the factors, both intrinsic and extrinsic, that affect animal movement, and ultimately how individual behaviour and movement scales to population-level processes. In the past 20 years, the ability to remotely track animals has revolutionised the field of movement ecology, especially via GPS (Global Positioning System) technol-

ogy (Rutz & Hays, 2009; Hussey *et al.*, 2015). To make sense of the corresponding new types of data, various statistical models were developed and are now routinely applied by ecologists (Morales *et al.*, 2004; Jonsen *et al.*, 2005; Johnson *et al.*, 2008; Patterson *et al.*, 2009). However, in recent years, we have witnessed a second wave of advancements in biologging technology, perhaps most notably accelerometry, which provide new opportunities but also pose new statistical challenges (Leos-Barajas *et al.*, 2017a). Generally, we are now able to remotely track and monitor animals at increasingly long time scales but at the same time also at increasingly fine temporal resolutions. To make full use of the corresponding complex new types of data, the ecologist's toolbox for analysing biologging data needs to be extended to include a modelling framework that is sufficiently versatile to accommodate heterogeneous, multi-stream, and multi-scale data — e.g. high frequency activity data via accelerometers and relatively low resolution geolocation data via active tracking or GPS tags — in a single model.

Over the last decade, hidden Markov models (HMMs) have emerged as a popular tool for animal movement modelling, where they provide a natural framework to infer behavioural modes and their various drivers from different types of telemetry data (Michelot *et al.*, 2016; Whoriskey *et al.*, 2017; Grecian *et al.*, 2018). A *basic* HMM comprises an observed movement process that depends on a *single* underlying hidden state process. The states are typically interpreted to be proxies for behavioural modes such as resting, foraging, or travelling. In ecological applications, it is of particular interest to make inference related to the influence of environmental covariates, e.g. regarding the behavioural response of blue whales (*Balaenoptera musculus*) to sonar exposure (DeRuiter *et al.*, 2016), the effect of wind speed on the flying dynamics of Verreaux's eagles (*Aquila verreauxii*; Leos-Barajas *et al.*, 2017a), or diel variation in Florida panther (*Puma concolor*) movements (Li & Bolker, 2017). In all these examples, HMMs have been proven to be a useful tool to reveal new insights into the interaction between animals and their environment.

While observations in a *basic* HMM are typically required to be regularly sampled

in time (or to follow some other regular sampling protocol, e.g. with observations made whenever a marine mammal comes to the surface to breathe), recent advances in biologging technology have led to a variety of novel telemetry sensors which often collect data from the same individual simultaneously at different time scales. Typical examples are locations in space and time obtained from GPS tags every hour, depth measurements obtained from time-depth recorders every ten minutes, or accelerations obtained from accelerometers several times per second. Since different types of behaviours may manifest themselves at different time scales (Leos-Barajas *et al.*, 2017b; Michelot *et al.*, 2017), being able to collect such multiple data streams, with differing temporal resolutions, offers various opportunities for ecological inference. Furthermore, by considering data collected at different time scales, joint models of such multi-scale data may contribute to reducing the effect of the often arbitrarily chosen time intervals between observations.

However, as the state process in a *basic* HMM operates on the same time scale as the movement process, these models do not readily accommodate multi-scale data. What usually would be done to model such data within an HMM framework is either to down-sample the observations from the different data streams to the *coarsest* of the different time scales (e.g. to process hourly observations into daily means of those observations, which can lead to a substantial loss of information that is actually contained in the raw data; c.f. Griffiths *et al.*, 2018), or to fit separate HMMs for the different variables, which conceptually is clearly inferior to formulating and fitting a joint model for the different variables, in particular with regard to identifying behavioural modes that affect multiple observed variables. We demonstrate that these problems can to some extent be overcome using *hierarchical* HMMs (HHMMs), where the observations are regarded as stemming from *multiple*, connected behavioural processes, each of which operates at the time scale at which the corresponding variables were observed.

HHMMs originate from the field of machine learning, where they have been introduced as a versatile tool for pattern recognition tasks. In handwriting or voice recognition,

for instance, different scales may be single letters or syllables, words, and sentences (Fine *et al.*, 1998). The hierarchy in that case results from the fact that several letters or syllables taken together constitute a word, several words taken together constitute a sentence, and so forth. Within HHMMs, these different levels are modelled using (distinct) hidden state processes, which are correlated with each other. In ecological applications, HHMMs have been previously applied in Leos-Barajas *et al.* (2017b), however considering only a *single* observed process, e.g. horizontal movements. Here we extend those models to allow for *multiple* observed processes, e.g. horizontal and vertical movements, or accelerations, which may be observed at different time scales. By incorporating multiple data streams, collected at different temporal resolutions, corresponding models may provide a much more comprehensive picture of animal movement with clear implications for ecological inference and conservation.

The paper is structured as follows: in Section 2, we introduce the main components of HHMMs, discuss the underlying dependence assumptions, and provide some details on maximum likelihood estimation of the model parameters. In addition, we briefly outline further topics related to HHMMs, including model selection, model checking, and state decoding. In Section 3, we illustrate the suggested approach in two real-data applications, where we model i) *coarse*-scale horizontal and *fine*-scale vertical Atlantic cod (*Gadus morhua*) movements throughout the English Channel, and ii) *coarse*-scale horizontal movements and corresponding *fine*-scale accelerations of a horn shark (*Heterodontus francisci*) tagged off the Californian coast.

2 Methodology

2.1 Hidden Markov models

A *basic* HMM for a P-dimensional time series comprises an *observed* state-dependent process $\{\mathbf{Y}_t\}_{t=1,\dots,T}$, $\mathbf{Y}_t = (Y_{1,t}, \dots, Y_{P,t})$ (e.g. hourly step lengths and turning angles, in

which case $P = 2$), which is driven by a *hidden* state process $\{S_t\}_{t=1,\ldots,T}$. The state process is typically modelled as a discrete-time, N-state Markov chain with transition probability matrix (t.p.m.) $\mathbf{\Gamma} = (\gamma_{ij})$, where

$$\gamma_{ij} = \Pr(S_t = j | S_{t-1} = i), \tag{1}$$

$i, j = 1, \ldots, N$, denotes the probability of switching from state i at time $t-1$ to state j at time t, and initial distribution vector $\boldsymbol{\delta} = (\delta_i)$, where

$$\delta_i = \Pr(S_1 = i), \tag{2}$$

$i = 1, \ldots, N$, denotes the probability of state i being active at time $t = 1$. Conditional on $S_t = i$, i.e. on state i being active at time t, the observation vector $\mathbf{Y}_t$ is drawn from a distribution associated with state i, defined by the P-dimensional probability (density) function $f^{(i)}(\mathbf{y}_t)$. Conditional on the entire state sequence, the observations $\mathbf{Y}_1, \ldots, \mathbf{Y}_T$ are assumed to be independent of each other. Furthermore, it is convenient to also assume the P different variables at time t to be conditionally independent of each other, given the state at time t, such that $f^{(i)}(\mathbf{y}_t)$ is the product of univariate densities (or probabilities, in the discrete case) $f_k^{(i)}(y_{k,t})$, $k = 1, \ldots, P$, i.e.

$$f^{(i)}(\mathbf{y}_t) = \prod_{k=1}^{P} f_k^{(i)}(y_{k,t}), \tag{3}$$

$i = 1, \ldots, N$. These dependence assumptions (i.e. Markov property, conditional independence across time and across variables) substantially facilitate statistical inference in HMMs, but can in certain scenarios be unrealistic and may then need to be relaxed. The dependence structure of such a basic HMM is illustrated in Figure 1.

Under the dependence assumptions made above, the likelihood of an HMM is given by

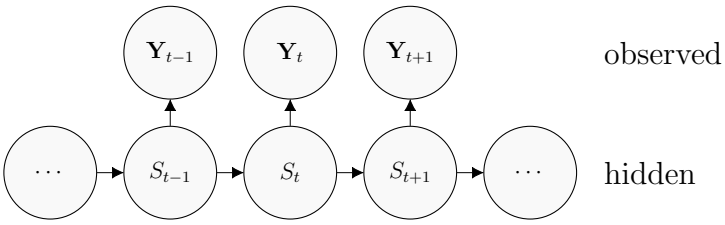

Figure 1: Dependence structure of a basic hidden Markov model.

the matrix product

$$\mathcal{L}^{\text{HMM}}(\mathbf{y}) = \boldsymbol{\delta}\mathbf{P}(\mathbf{y}_1)\prod_{t=2}^{T}\boldsymbol{\Gamma}\mathbf{P}(\mathbf{y}_t)\mathbf{1}, \tag{4}$$

with diagonal matrix $\mathbf{P}(\mathbf{y}_t) = \text{diag}\big(f^{(1)}(\mathbf{y}_t), \ldots, f^{(N)}(\mathbf{y}_t)\big)$, and $\mathbf{1} \in \mathbb{R}^N$ is a column vector of ones. Evaluation of (4) corresponds to applying the so-called forward algorithm, which is a powerful tool that renders likelihood-based inference in HMMs fast and convenient; see Zucchini *et al.* (2016) for details.

2.2 Hierarchical hidden Markov models

To extend the basic HMM in a way that allows for joint inference at multiple time scales, we first distinguish between an *observed coarse*-scale P-dimensional state-dependent process $\{\mathbf{Y}_t\}_{t=1,\ldots,T}$ (e.g. hourly step lengths and turning angles), driven by a *hidden coarse*-scale state process $\{S_t\}_{t=1,\ldots,T}$, and an *observed fine*-scale P^*-dimensional state-dependent process $\{\mathbf{Y}^*_{t,t^*}\}_{t^*=1,\ldots,T^*}$ (e.g. accelerations observed at say 1 hertz, in one, two, or three dimensions), driven by a *hidden fine*-scale state process $\{S^*_{t,t^*}\}_{t^*=1,\ldots,T^*}$; see Figure 2 for an illustration of the model structure. We then segment the *fine*-scale observations into T distinct chunks, each of length T^*, such that each chunk contains all *fine*-scale observations that were observed during the t-th sampling of the *coarse*-scale state process (e.g. all $T^* = 3600$ accelerations that were observed during the t-th sampling of hourly step lengths and turning angles). Subsequently, we connect each chunk of *fine*-scale observations to one of N possible *fine*-scale HMMs, each of which has its own t.p.m. $\boldsymbol{\Gamma}^{*(i)} = (\gamma^{*(i)}_{k^*l^*})$, initial

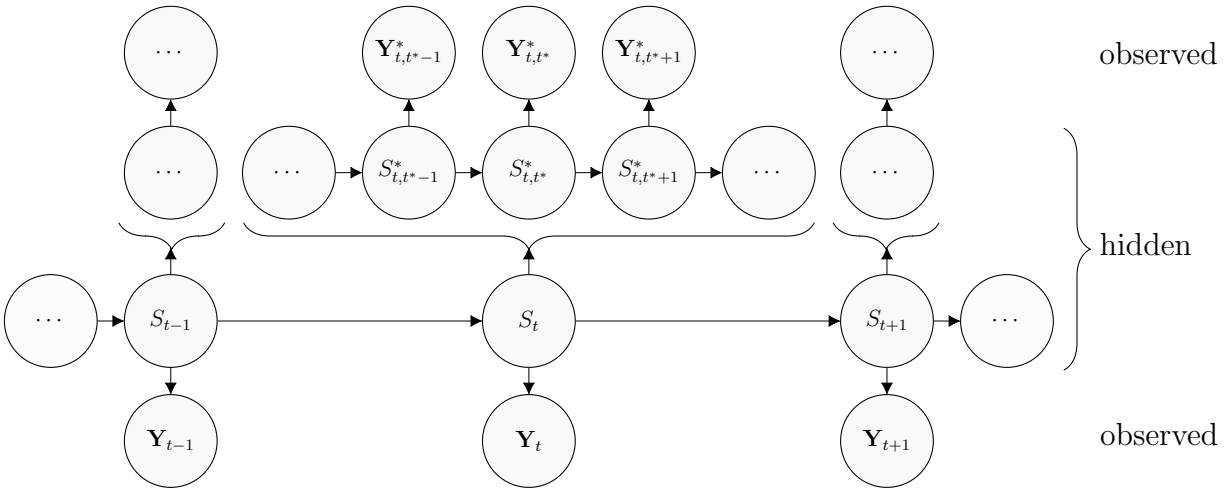

Figure 2: Dependence structure of a basic hierarchical hidden Markov model.

distribution $\boldsymbol{\delta}^{*(i)} = (\delta^{*(i)}_{k^*})$, and state-dependent distributions $f^{*(i,k^*)}(\mathbf{y}^*_{t,t^*})$, $i = 1, \ldots, N$, $k^*, l^* = 1, \ldots, N^*$. The state of the *coarse*-scale state process that is active at time t, $S_t = i$, is now assumed to select both the state-dependent distribution which generates the observations at the *coarse* scale as well as the HMM which generates all *fine*-scale observations during the t-th sampling of the *coarse*-scale state process. Thus, there is one HMM (with its own t.p.m. and state-dependent distributions) associated with each of the *coarse*-scale states.

We assume both state processes to be of first order (Markov property), and both state-dependent processes to satisfy the two conditional dependence assumptions (across time and across variables) as detailed in Section 2.1. The two different state processes can be thought of as proxies for behavioural modes, or movement strategies, relevant at longer term (*coarse*-scale state process) and shorter term (*fine*-scale state process) scales, respectively.

Similar to (4), the likelihood of an HHMM is obtained as the matrix product

$$\mathcal{L}^{\text{HHMM}}(\mathbf{y}, \mathbf{y}^*) = \boldsymbol{\delta}\mathbf{P}(\mathbf{y}_1, \mathbf{y}^*_1) \prod_{t=2}^{T} \boldsymbol{\Gamma}\mathbf{P}(\mathbf{y}_t, \mathbf{y}^*_t)\mathbf{1}, \tag{5}$$

with diagonal matrix $\mathbf{P}(\mathbf{y}_t, \mathbf{y}_t^*) = \text{diag}\big(\mathcal{L}_1^{\text{HMM}^*}(\mathbf{y}_t^*) f^{(1)}(\mathbf{y}_t), \ldots, \mathcal{L}_N^{\text{HMM}^*}(\mathbf{y}_t^*) f^{(N)}(\mathbf{y}_t)\big)$, where $\mathcal{L}_i^{\text{HMM}^*}(\mathbf{y}_t^*)$ is the likelihood of the t-th chunk of *fine*-scale observations being generated by the i-th *fine*-scale HMM. A recursive algorithm to efficiently evaluate (5), which renders a numerical maximisation of the likelihood fast and convenient, is provided in the Appendix.

2.3 Incorporating covariates

Covariates can be included in HHMMs by expressing some of the model parameters as functions of these covariates. In principle, covariates can be included both in the state-dependent processes, where they determine the parameters of the state-dependent distributions, and in the state processes, where they affect the transition probabilities. Here we focus on the latter as in ecological applications the interest usually lies in modelling the effect of covariates on state occupancy.

To incorporate covariates in the *coarse*-scale state process, we express the transition probabilities γ_{ij} at time t as a function of some predictor $\eta_{ij}(\mathbf{x}_t)$, which in turn depends on a Q-dimensional covariate vector $\mathbf{x}_t = (x_{1,t}, \ldots, x_{Q,t})$. Using the multinomial logit link to ensure that $0 < \gamma_{ij}(\mathbf{x}_t) < 1$ and $\sum_{j=1}^{N} \gamma_{ij}(\mathbf{x}_t) = 1$ for all i, we obtain the t.p.m. $\mathbf{\Gamma}(\mathbf{x}_t) = \big(\gamma_{ij}(\mathbf{x}_t)\big)$, with

$$\gamma_{ij}(\mathbf{x}_t) = \frac{\exp(\eta_{ij}\big(\mathbf{x}_t)\big)}{\sum_{k=1}^{N} \exp\big(\eta_{ik}(\mathbf{x}_t)\big)}, \text{ where } \eta_{ij}(\mathbf{x}_t) = \begin{cases} \beta_0^{(ij)} + \sum_{k=1}^{Q} \beta_k^{(ij)} x_{k,t} & \text{if } i \neq j; \\ 0 & \text{otherwise.} \end{cases} \tag{6}$$

Instead of estimating the state transition probabilities directly, we then estimate the regression coefficients $\beta_k^{(ij)}$, $i, j = 1 \ldots, N$, $i \neq j$, $k = 0, \ldots, Q$.

In some applications, it is of particular interest to model seasonal variation or within-day variation, i.e. to consider the time as a deterministic covariate rather than as a stochastic one. To account for the corresponding periodic effects, trigonometric functions can be

used, where the predictor in (6) can be rewritten as

$$\eta_{ij}(\mathbf{x}_t) = \begin{cases} \beta_0^{(ij)} + \beta_1^{(ij)} \sin\left(\frac{2\pi t}{r}\right) + \beta_2^{(ij)} \cos\left(\frac{2\pi t}{r}\right) & \text{if } i \neq j; \\ 0 & \text{otherwise,} \end{cases} \tag{7}$$

where r denotes the length of the period of interest (e.g. $r = 24$ in case of within-day variation and hourly observations). For more flexibility, additional sine and cosine terms with shorter cycles can be added to the predictor.

Covariate inclusion in the *fine*-scale state-process is analogous, though note that in this case there is one t.p.m. $\mathbf{\Gamma}^{*(i)}(\mathbf{x}_t) = \left(\gamma_{k^*l^*}^{*(i)}(\mathbf{x}_t)\right)$ for each state i of the *coarse*-scale state process, i.e. N such matrices to be expressed as functions of covariates.

2.4 Model fitting

Using the forward algorithm, evaluation of (5) requires $\mathcal{O}(NT^*N^{*2} + TN^2)$ operations, which renders a numerical maximisation of the likelihood using some Newton-Raphson-type optimisation routine, e.g. the R function nlm (R Core Team, 2017), practically feasible even for long time series and moderate numbers of states. Typical challenges related to numerical likelihood maximisation in HMMs, particularly parameter constraints, numerical underflow, and local maxima, also apply to HHMMs. Specifically, to account for parameter constraints, we transform the constrained parameters into unconstrained parameters using some one-to-one transformation and maximise the likelihood with respect to the unconstrained parameters. To avoid numerical underflow, which may result from multiplying a large number of small probabilities in the likelihood calculation, we maximise the log-likelihood and evaluate all quantities on the log-scale; see the implementation of the forward algorithm in the Appendix. As the numerical maximisation might yield a local rather then the global maximum of the likelihood, using good initial values for the search is crucial. To increase the chance of having found the global maximum, a good strategy is

to run the search from a range of different, possibly randomly selected initial values and select the model that corresponds to the largest likelihood.

2.5 Model selection

Model selection primarily involves the specification of the state-dependent distributions, the selection of the number of states (order selection), and variable selection (in case of covariates being included), but could extend also to investigations of possible assumption violations particularly regarding the dependence structure. The state-dependent distributions are typically determined by the data type of the variables considered: for positive continuous-valued variables (e.g. step lengths, dive depths, or vertical movements), for instance, gamma distributions provide a natural choice, whereas for circular variables (e.g. turning angles), von Mises or wrapped Cauchy distributions are commonly used. Information criteria, such as Akaike's information criterion or the Bayesian information criterion, provide a natural approach to order selection in HHMMs when fitted in a maximum likelihood framework. However, it has been shown that these criteria tend to favor overly complex HMMs, with more states than seem biologically plausible, when used for ecological data with complex features (Li & Bolker, 2017; Pohle *et al.*, 2017). As these practical problems are likely exacerbated in the more complex HHMM framework, we strongly advise against reliance on such criteria. Instead, we recommend to closely inspect fitted models with different numbers of states, and then pragmatically choose the numbers of states at both levels taking into account (ecological) expert knowledge.

2.6 Model checking

Model checking in HMM-type models is typically done based on so-called pseudo-residuals, which use the probability integral transformation to assess whether any given observation is well explained by the fitted model. For the *coarse*-scale observations, evaluation of the

pseudo-residuals proceeds as in basic HMMs (see Zucchini *et al.*, 2016). For the *fine*-scale observations, it is convenient to first decode the *coarse*-scale Markov chain (see subsequent section) and then to compute the pseudo-residuals separately for each chunk of *fine*-scale observations conditional on the HMM active according to the decoded *coarse*-scale state. Other strategies for model checking include comparing the empirical distribution of any variable observed to the corresponding marginal distribution as implied under the fitted model, or simulations from the fitted model to check if it can reproduce the key patterns found in the empirical data (c.f. Langrock *et al.*, 2013).

2.7 State decoding

In ecological applications, it is often of particular interest to decode the hidden states, i.e. to compute the most likely sequence of states that may have given rise to the observations under the fitted model. The simplest and most natural approach is to first decode the *coarse*-scale states $s_1, \ldots, s_T$ (taking into account both *coarse*-scale and *fine*-scale observations), and then, for any time t of the *coarse*-scale state process, decode the *fine*-scale states $s^*_{t,1}, \ldots, s^*_{t,T^*}$ conditional on the most likely *coarse*-scale state to be active at time t (taking into account only the *fine*-scale observations). The decoding can be done either locally, considering each time point in isolation, or globally, considering the time series as a whole. In practice, global decoding, which can conveniently be carried out using the Viterbi algorithm, is usually the default choice and therefore used throughout this paper. Technical details on both local and global decoding transfer directly from basic HMMs to HHMMs; see Zucchini *et al.* (2016) for details.

3 Real-data applications

3.1 Atlantic cod movement

Atlantic cod (*Gadus morhua*) is a commercially valuable demersal fish species found throughout the shelf seas surrounding the British Isles. For decades, cod have been heavily exploited, resulting in steep declines in abundance. Despite significant attention to stock assessment and fisheries management, efforts to rebuild cod stocks have had limited success (ICES, 2018). To facilitate informed conservation actions, information about when, where, and how individuals move and undertake key life-history events are essential (Hays *et al.*, 2019; Hussey *et al.*, 2015). The tagging of cod to gain these information has been ongoing for a number of years and has greatly increased our knowledge of individual behaviour (Righton *et al.*, 2001; Hobson *et al.*, 2007; Hobson *et al.*, 2009) and stock structure (Neat *et al.*, 2014).

To illustrate the application of HHMMs, we jointly model *coarse*-scale horizontal and *fine*-scale vertical movements from a single cod, where we are particularly interested in understanding diel and circatidal patterns in the cod's vertical movements and how the latter are driven by its horizontal movements. As demersal fish rarely swim in surface waters (which is a pre-requisite for satellite tags; Rutz & Hays, 2009), tagging was achieved using an archival data storage tag (DST). DSTs are typically pre-programmed to record depth at regular time intervals for the duration of deployment (here every 10 minutes). From this, we calculated log-vertical movements (log-differences in depth per 10-minute interval in meters) and estimated daily geolocations using a single-state version of the tidal geolocation model of Pedersen *et al.* (2008), which were then processed to give daily step lengths and turning angles. The method has been adapted to ensure that the underlying diffusion model operates under a fixed diffusivity parameter (30 kilometers per day^2) and does not switch between two based on the presence or absence of a tidal signal. As some (one or two per day) of each day's 144 depth observations are being used to generate

the cod's geolocations, some minor conditional dependence between the two movement rates is expected, which for simplicity is neglected in the model formulation. To clarify, only one (no tidal signal) or two (during the tidal signal and not during the tidal signal) depth observations representing the daily maxima are used per day. This is because they — alongside model-fitted estimates of tidal range and time — have shown to yield accurate geolocations in a number of demersal fish species, including, *inter alia*, European plaice (*Pleuronectes platessa*; Hunter *et al.*, 2004), thornback ray (*Raja clavata*; Hunter *et al.*, 2005), and Atlantic cod (Righton *et al.*, 2007). For further details we direct the interested reader to Pedersen *et al.* (2008). Thus, we ended up with two separate time series sampled at different time scales: vertical movements at 10-minute intervals and horizontal movements at daily intervals, i.e. for each of the $T = 291$ daily horizontal movements, we have $T^* = 144$ 10-minute vertical movements. Previous work has often overcome this difference in sampling by either gaining meaningful inference from a single dimension (e.g. Hobson *et al.*, 2007) or, in the case of Griffiths *et al.* (2018), who analyse movements in both dimensions, by simplifying the vertical dimension at the daily scale. Here we demonstrate how HHMMs can be used to analyse movements in both dimensions while retaining the vertical dimension at the 10-minute scale.

Based on an exploratory data analysis and a comparison of fitted models with different numbers of states, we chose $N = 3$ states for the *coarse*-scale state process, as visual inspection of the data revealed two different types of horizontal movements, one of which corresponds to again two different vertical movement patterns, which can only be captured if a third state is considered at the *coarse* scale. Each of the *coarse*-scale states was then associated with an HMM with $N^* = 3$ *fine*-scale states (resulting in 9 *fine*-scale states in total), which allows us to draw a nuanced yet not overly complex picture of the cod's vertical movements. To model diel variation in the vertical movements, the transition probabilities of the Markov chains determining the *fine*-scale state processes were estimated as functions of the time of day, with the predictors specified as given

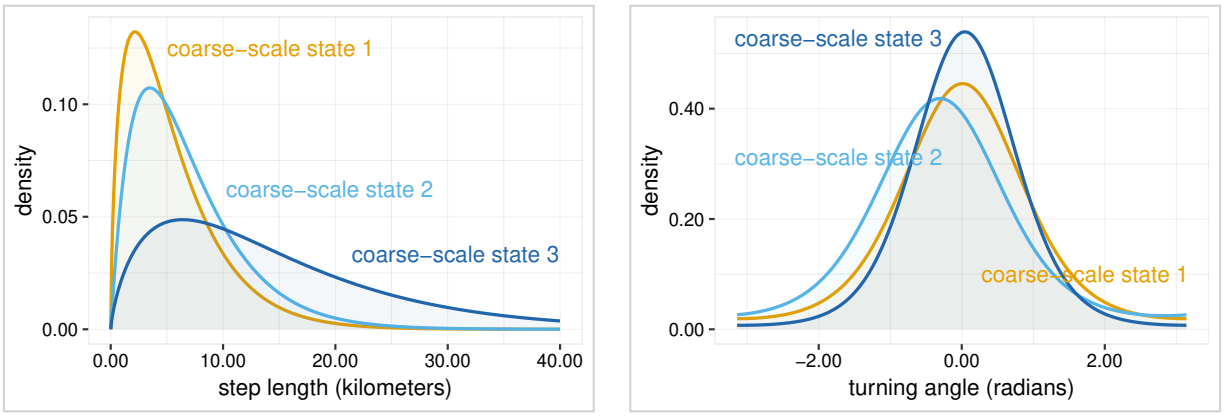

Figure 3: Estimated state-dependent distributions of *coarse*-scale step lengths (left panel) and turning angles (right panel) of an Atlantic cod. *Coarse*-scale states 1 and 2 may be interpreted as resting/foraging and mobile/foraging behaviours, respectively, which are very similar in terms of horizontal movements but substantially differ in the corresponding vertical movement patterns, while *coarse*-scale state 3 may be linked to a travelling/migrating behaviour.

in (7); the *coarse*-scale state transition probabilities were assumed to be constant over time. For the step lengths and vertical movements, we assumed gamma distributions (with an additional point mass on zero in case of the vertical movements to account for the zeros observed), while for the turning angles, von Mises distributions were considered. The computation time required to fit the model was 6.1 hours, where the likelihood was evaluated in C++ and numerically maximised using the R function nlm (R Core Team, 2017) on a 3.6 GHz Intel® Core™ i7 CPU.

The estimated state-dependent distributions of *coarse*-scale step lengths and turning angles are displayed in Figure 3. *Coarse*-scale states 1 and 2 capture short, slightly less directed horizontal movements, where we interpret *coarse*-scale state 1 as a resident/foraging behaviour and *coarse*-scale state 2 as a mobile/foraging behaviour. Although these two states are very similar in terms of horizontal movements, they substantially differ in the corresponding vertical movement patterns (see below). *Coarse*-scale state 3 relates to relatively longer, slightly more directed horizontal movements, which may be linked to a travelling/migrating behaviour. The t.p.m. of the *coarse*-scale state process was estimated

as

$$\hat{\mathbf{\Gamma}} = \begin{pmatrix} 0.945 & 0.000 & 0.055 \\ 0.064 & 0.777 & 0.160 \\ 0.098 & 0.075 & 0.827 \end{pmatrix},$$

which implies the stationary distribution $(0.618, 0.096, 0.286)$, indicating that approximately 62% (180 days), 10% (28 days), and 29% (83 days) of the observations were generated in *coarse*-scale state 1, 2, and 3, respectively.

The estimated state-dependent distributions of *fine*-scale vertical movements and the associated stationary distributions of the corresponding *fine*-scale state processes as functions of the time of day along with 95% confidence intervals are displayed in Figure 4. When the cod was in *coarse*-scale state 1 (resting/foraging), then the vertical movements were generated by the three state-dependent distributions displayed in the upper panel. The level of vertical movement was fairly *low* (according to the stationary distribution, the cod was in *fine*-scale state 3, which corresponds to a *relatively* high level of vertical movement, less than 20% of the time) and slightly *increased* during the day (where it spent up to 75% of the time in *fine*-scale state 2, which corresponds to a *relatively* moderate level of vertical movement, and less than 25% of the time in *fine*-scale state 1, which corresponds to a *relatively* low level of vertical movement). When the cod was in *coarse*-scale state 2 (mobile/foraging), then the vertical movements were generated by the three state-dependent distributions displayed in the center, which correspond to a very *low* level of vertical movement (all three state-dependent distributions have considerably smaller means than those corresponding to *coarse*-scale states 1 and 3), where state occupancy (as indicated by the associated stationary distribution) is not much affected by the time of day. When the cod was in *coarse*-scale state 3 (travelling/migrating), then the vertical movements were generated by the three state-dependent distributions displayed in the lower panel. Here the opposite was the case: the level of vertical movement was much *higher* relative to *coarse*-scale states 1 and 2 (*fine*-scale state 3, whose state-dependent

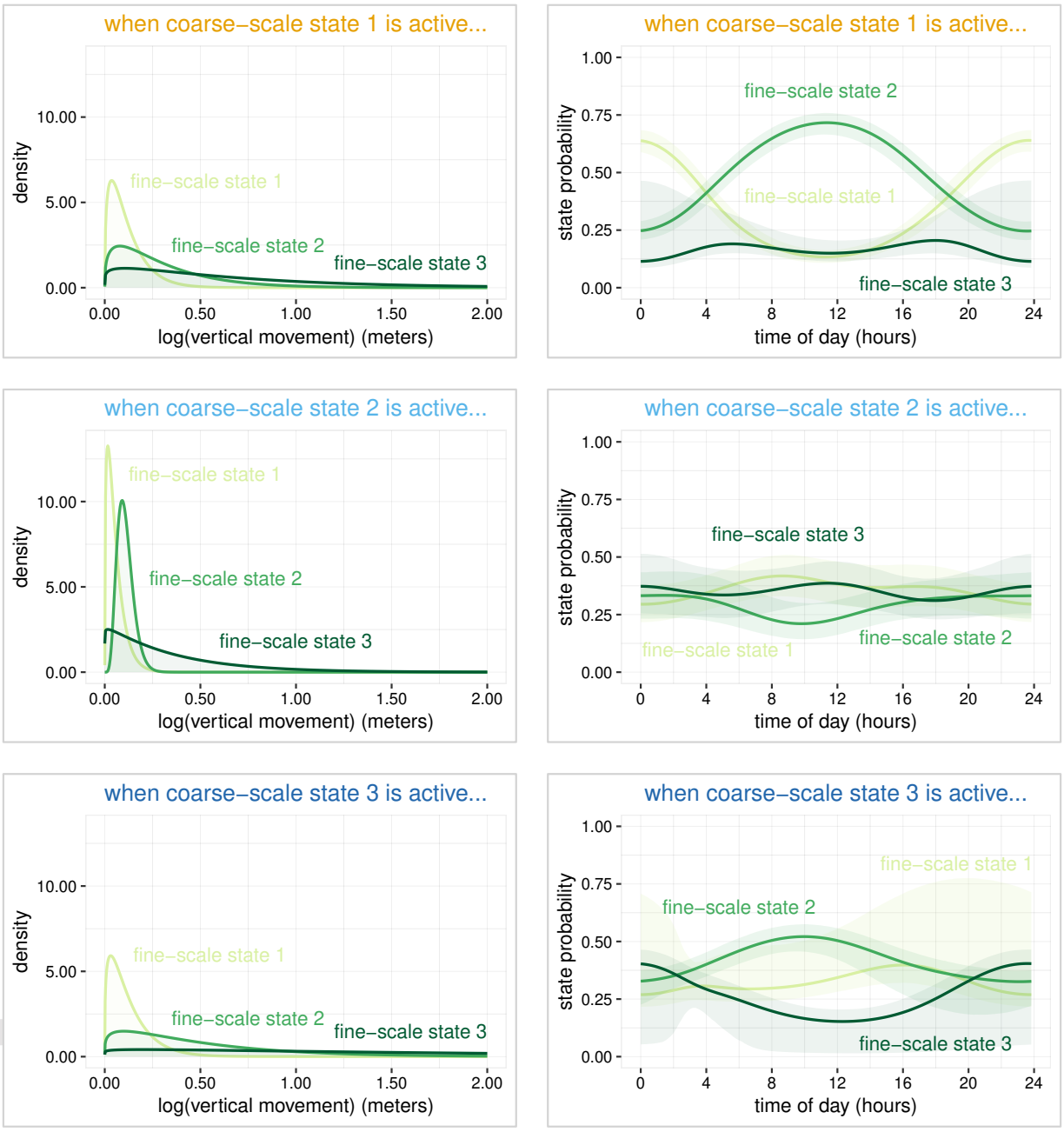

Figure 4: Estimated state-dependent distributions of *fine*-scale vertical movements (left panel) of an Atlantic cod and associated stationary distributions of the corresponding *fine*-scale state processes as functions of the time of day along with 95% confidence intervals (right panel) for *coarse*-scale state 1 (resting/foraging), 2 (mobile/foraging), and 3 (travelling/migrating), respectively. *Fine*-scale states 1, 2, and 3 represent *relatively* low, moderate, and high levels of vertical movement, respectively, where the corresponding levels substantially differ across the different *coarse*-scale states (the means of the state-dependent distributions for *fine*-scale state 3, for instance, vary from 0.355 in *coarse*-scale state 2 over 0.689 in *coarse*-scale state 1 to 1.983 in *coarse*-scale state 3).

distribution has a mean of 1.983 and therefore captures much higher vertical movement activity than those corresponding to *fine*-scale state 3 within the HMMs corresponding to *coarse*-scale states 1 (0.689) and 2 (0.355), was — depending on the time of day — active between 15% and 45% of the time) and slightly *decreased* during the day (though note that due to the relatively high uncertainty associated with the stationary distributions these results should be treated with some caution).

The decoded horizontal movement track and three example sequences of *fine*-scale vertical movements for the different *coarse*-scale states are displayed in Figure 5, where the decoding was performed using the Viterbi algorithm as described in Section 2.7. The cod spent most of its time (178 days) in *coarse*-scale state 1, where reduced rates of movement indicate prolonged periods of resting or localised foraging. This was then interspersed by two travelling/migrating periods associated with *coarse*-scale state 3 (81 days) as the cod traversed the English Channel, and some periods spent in *coarse*-scale state 2 (32 days).

Throughout the time spent in *coarse*-scale state 1, the associated *fine*-scale state process exhibited clear diurnal patterns (similar trends are found in Løkkeborg, 1998). During the day, the level of vertical movement increased, as the cod was more likely to switch from *fine*-scale state 1 to *fine*-scale state 2. This may be interpreted as localised foraging, as cod frequently move off the seafloor to pursue benthic-dwelling prey (e.g. crustaceans or other small fish species) via visual predation (Adlerstein & Welleman, 2000; Hobson *et al.*, 2009). The increased probability of switching back to *fine*-scale state 1 during the night points towards a much more resting-like behaviour as the cod returns to the seafloor (as observed in Hobson *et al.*, 2007). *coarse*-scale state 2, in comparison, involves a much lower level of vertical movement, which is not much affected by the time of day. This could indicate an intermediate behavioural type, where the cod was foraging and remained close to the seabed while being slightly more mobile in the horizontal dimension relative to *coarse*-scale state 1 (i.e. it was not resident). Throughout the time spent in *coarse*-scale state 3, the cod was clearly migrating, exhibiting increased rates of horizontal movement,

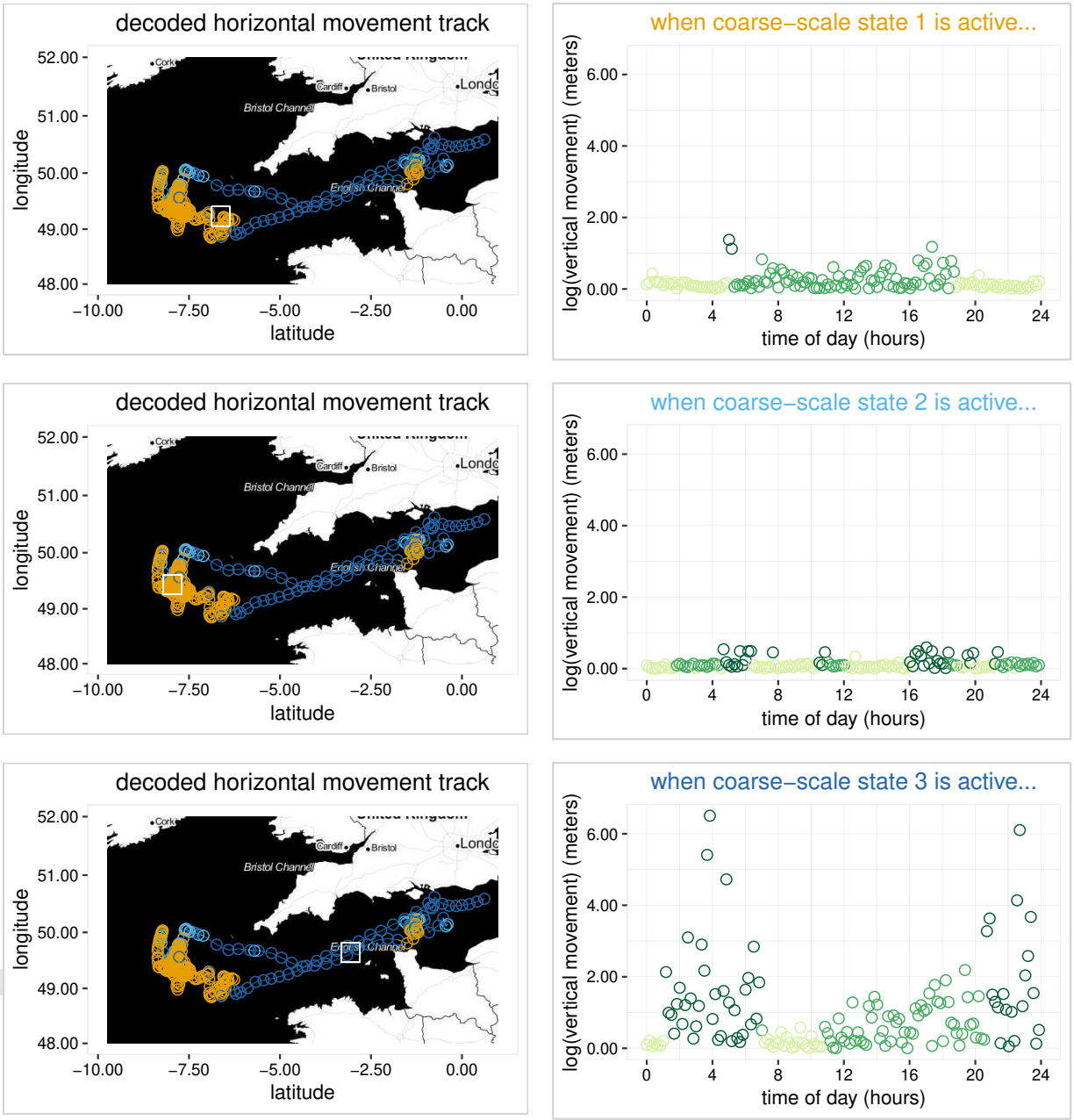

Figure 5: Decoded horizontal movement track (left panel) and three example sequences of decoded *fine*-scale vertical movements (right panel) of an Atlantic cod, one for *coarse*-scale state 1 (resting/foraging; $t = 252$), one for *coarse*-scale state 2 (mobile/foraging; $t = 86$), and one for *coarse*-scale state 3 (travelling/migrating; $t = 48$), each of length 24 hours. The white squares on the maps in the left panel indicate the days which were chosen for the three example sequences displayed in the right panel.

slightly more uniform directionality, and elevated rates of vertical movement as it transits the English Channel. Greater vertical movements during migration could indicate the use of circatidal selective tidal stream transport (STST), as the cod moves up off the seabed into the water column during favourable tides and uses the tide's velocity to efficiently cruise in the desired direction. STST is more commonly seen in flatfish such as European plaice (*Pleuronectes platessa*; Hunter *et al.*, 2004), however, cod have been shown to use this highly efficient means of transport during migration in the North Sea (Righton *et al.*, 2007).

Two findings are noteworthy: first, the presence of diel vertical movement patterns in the *fine*-scale state process associated with *coarse*-scale state 3 (travelling/migrating), which illustrates that vertical activity is high during the night (as illustrated in the bottom right panel of Figure 5); second, that *coarse*-scale state 2 (mobile/foraging) mostly occurs during post-spawning migration, as cod transit from spawning grounds in the southern North Sea to feeding grounds in the eastern English Channel. The *fine*-scale patterns of vertical movement identified during *coarse*-scale state 1 are clear indications that cod are able to vary their feeding and foraging patterns in relation to prey availability, whether prey are available by day, by night, or only during crepuscular periods. The variable patterns of vertical movement across *coarse*-scale states 2 and 3 may suggest that cod are capable of migrating quickly to reach spawning grounds after a summer spent foraging on rich feeding grounds or moving more slowly and taking advantage of food resources to recover energy after the spawning period. Such adaptive migratory behaviour could easily be overlooked by studies that consider movement in only one dimension (e.g. Hobson *et al.*, 2007) or studies that only consider movement rates at the daily scale (e.g. Griffiths *et al.*, 2018), which highlights the potential of the HHMM approach.

Quantile-quantile plots and autocorrelation functions of normal pseudo-residuals for *coarse*-scale step lengths and turning angles and three example sequences of *fine*-scale vertical movements, each computed as described in Section 2.6, are displayed in Figure

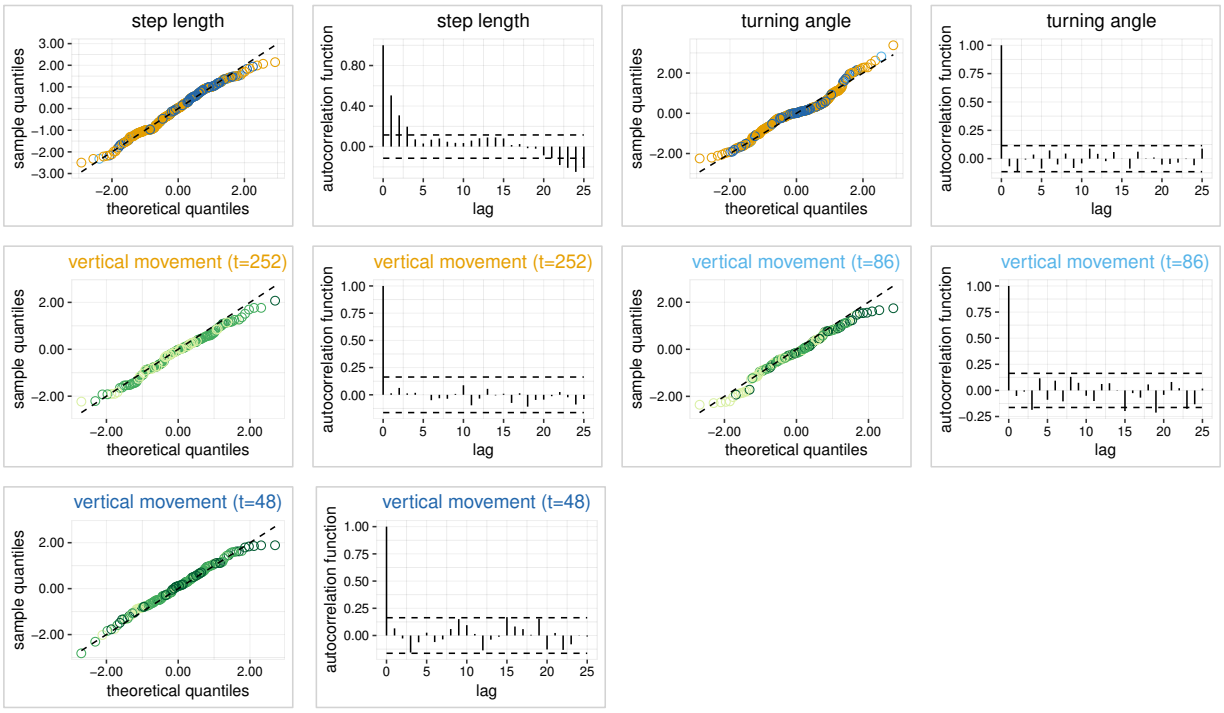

Figure 6: Quantile-quantile plots and autocorrelation functions of normal pseudo-residuals for *coarse*-scale step lengths and turning angles and three example sequences of *fine*-scale vertical movements, one for *coarse*-scale state 1 (resting/foraging; $t = 252$), one for *coarse*-scale state 2 (mobile/foraging; $t = 86$), and one for *coarse*-scale state 3 (travelling/migrating; $t = 48$; each of length 24 hours). The three example sequences considered here are the same as those displayed in Figure 5, where the colours indicate the Viterbi-decoded states.

6. The plots indicate some minor lack of fit regarding the marginal distributions of the different variables, and also some residual correlation in the step lengths series. Overall, the magnitude of the lack of fit found here is anything but unusual for movement modelling exercises, due to the very complex patterns typically found in such data. Thus, we consider the goodness of fit of our model to be satisfactory. In principle, more flexible state-dependent distributions can be used to improve the fit (c.f. Langrock *et al.*, 2018), which however we refrain from investigating further as our aim here is to present an illustrative case study (thus trading some relatively minor lack of fit against a more complex model formulation, which would complicate the interpretation).

3.2 Horn shark movement

Horn sharks (*Heterodontus francisci*) are abundant, nocturnally active sharks associated with rocky reefs and kelp beds in southern California (Nelson & Johnson, 1970; Love, 1991). Horn sharks are relatively small (< 122 centimeters) and as non-obligate ram ventilators they can rest on the sea floor for relatively long periods of time (Compagno, 2001). Despite their importance as meso-predators in rocky reef habitats, surprisingly little work has focused on their movements and behaviours (Strong, 1989; Compagno, 2001). Previous observations have found horn sharks to have a relatively high site fidelity to a reef (Strong, 1989). However, to what spatio-temporal extent they use the reef for various behaviours (i.e. resting, foraging, etc.) is unknown.

2 sources of data were collected from a horn shark at Santa Catalina Island, California, USA: geopositional and acceleration data. Custom tag packages were externally attached to the dorsal fin of the shark that included an accelerometer data logger (Technosmart, AxyDepth) and an acoustic transmitter (Vemco V9 continuous pulse pinger, 2000 milliseconds). The shark was actively tracked from a tracking vessel above using a Vemco VR100 on-board receiver and VH110 directional hydrophone. At the end of the tracking duration, divers relocated the shark to remove the tag package and download data from the accelerometer data logger. During active tracking, geopositions were estimated for the tracking vessel relative to that of the tagged shark. Although boat position was recorded every 2 seconds, position data were filtered according to the closest possible estimations of the shark to the boat. Geopositional error for the shark was estimated to be 5 metres (via range tests at the study location). However, at times the shark could be active yet simply not travel a large enough distance relative to positional measurement error to detect a change in geoposition. This leads to many consecutive data points with the same coordinates that, on their own, provide no indication of whether the shark was resting or not. Acceleration data were recorded continuously at 25 hertz and provide a much more fine-scaled picture of the overall body movements of the shark because it records move-

ments in three orthogonal dimensions, the dorso–ventral, anterior–posterior, and lateral axes (Shepard *et al.*, 2008). We summarised the acceleration signal across the 3 dimensions by calculating the overall dynamic body acceleration (ODBA) as a general measure of activity (Wilson *et al.*, 2006).

While acceleration data were continuously recorded, geoposition data were collected in bouts over the course of one night, resulting in $T = 2694$ geopositions observed at a temporal resolution of 2 seconds splitted into 194 segments, where the minimum (mean, maximum) number of observations in each segment is 6 (15, 96), corresponding to a duration of 12 (30, 192) seconds, respectively. As a general measure of displacement, we calculated the Euclidean distance, y_t, between consecutive geopositions, and paired each distance with the corresponding $T^* = 50$ ODBA values, y^*_{t,t^*}, produced during this time frame. Once we combine the two time series, we have two processes that we model: one occurring at a *coarse* scale of 2 seconds, producing a value of distance and 50 sequential ODBA values, and within the 2 seconds time frame, a *fine*-scale process that describes the evolution of the 50 ODBA values during the 2 seconds. Due to the measurement error associated, the horn shark may be active yet not travel far enough to record a change in geoposition. Keeping this in mind, we constructed $N = 3$ *coarse*-scale state processes that we generally interpret as i) zero distance travelled, $y_t = 0$, and containing relatively larger values of ODBA produced, indicating high activity, ii) zero distance travelled, $y_t = 0$, and mostly lower values of ODBA produced, indicating a resting period, and iii) some distance travelled, $y_t > 0$, and corresponding ODBA values. For $y_t > 0$, we defined categories of distances with cutoffs $y_0 < y_1 < \cdots < y_J$ in order to construct a so-called histogram distribution of step lengths. Overall, we define $N = 3$ *coarse*-scale state processes, each of which was then associated with an HMM with $N^* = 3$ *fine*-scale states. The specification of $N^* = 3$ for each *coarse*-scale state reflects that periods of active movement are a composition of three general activity levels, which, for consistency, we adopted when there is no change in geoposition. In this case study, no covariate influence was modelled. The

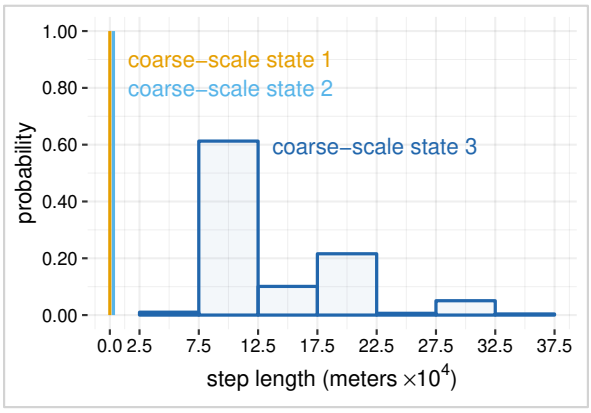

Figure 7: Estimated state-dependent histogram distribution of *coarse*-scale step lengths of a horn shark. *Coarse*-scale state 3 is indicative of travelling, while *coarse*-scale states 1 and 2 correspond to step lengths of zero, which substantially differ in the corresponding acceleration patterns.

computation time required to fit the model was 2.2 hours.

The estimated state-dependent distributions of *coarse*-scale step lengths are displayed in Figure 7. The t.p.m. of the *coarse*-scale state process was estimated as

$$\hat{\mathbf{\Gamma}} = \begin{pmatrix} 0.418 & 0.218 & 0.364 \\ 0.081 & 0.536 & 0.383 \\ 0.119 & 0.279 & 0.602 \end{pmatrix},$$

which implies the stationary distribution $(0.150, 0.363, 0.487)$, indicating that approximately 15% (798 seconds), 36% (1954 seconds), and 49% (2596 seconds) of the observations were generated in *coarse*-scale state 1, 2, and 3, respectively.

The estimated state-dependent distributions of *fine*-scale accelerations for the different *coarse*-scale states are displayed in Figure 8. The t.p.m.s of the *fine*-scale state processes

were estimated as

$$\hat{\mathbf{\Gamma}}^{*(1)} = \begin{pmatrix} 0.949 & 0.051 & 0.000 \\ 0.045 & 0.944 & 0.011 \\ 0.000 & 0.029 & 0.971 \end{pmatrix}, \hat{\mathbf{\Gamma}}^{*(2)} = \begin{pmatrix} 0.962 & 0.038 & 0.000 \\ 0.040 & 0.906 & 0.054 \\ 0.000 & 0.067 & 0.933 \end{pmatrix},$$

$$\hat{\mathbf{\Gamma}}^{*(3)} = \begin{pmatrix} 0.961 & 0.039 & 0.000 \\ 0.030 & 0.954 & 0.016 \\ 0.000 & 0.041 & 0.959 \end{pmatrix},$$

which imply the stationary distributions $(0.397, 0.444, 0.159)$, $(0.369, 0.348, 0.282)$ and $(0.360, 0.460, 0.179)$.

Coarse-scale states 1 and 2 are characterized by no change in estimated geoposition of the shark but, as seen in Figure 8, contrasting distributions of ODBA values were produced during the 2 second time frame. *Coarse*-scale state 1 generally produces much higher values of ODBA than *coarse*-scale state 2, indicating that the shark is quite active during these periods. Within *coarse*-scale state 1, the ODBA values captured in *fine*-scale state 3 can even be larger than the ODBA values produced when the shark is travelling (see also Figure 9 for example sequences where this is the case). While persistence in *coarse*-scale state 1 is low, as demonstrated by the t.p.m. of the *coarse*-scale state process, the track analysed is only from one night and for one shark. Overall, the shark spent about 15% of the time it was tracked not travelling but active, 36% of the time not moving and resting, and 49% of the time travelling around reefs at Santa Catalina Island.

This case study demonstrates how horn sharks can periodically rest in-between travelling bouts, making it difficult to distinguish between active and resting behaviours using conventional movement models from active tracking. The framework presented here demonstrates the beginning stages of combining multiple data streams to make inferences about the shark's behaviours, particularly when travel distance is zero. The horn shark in this example switched between behavioural states often along its traveled path, and

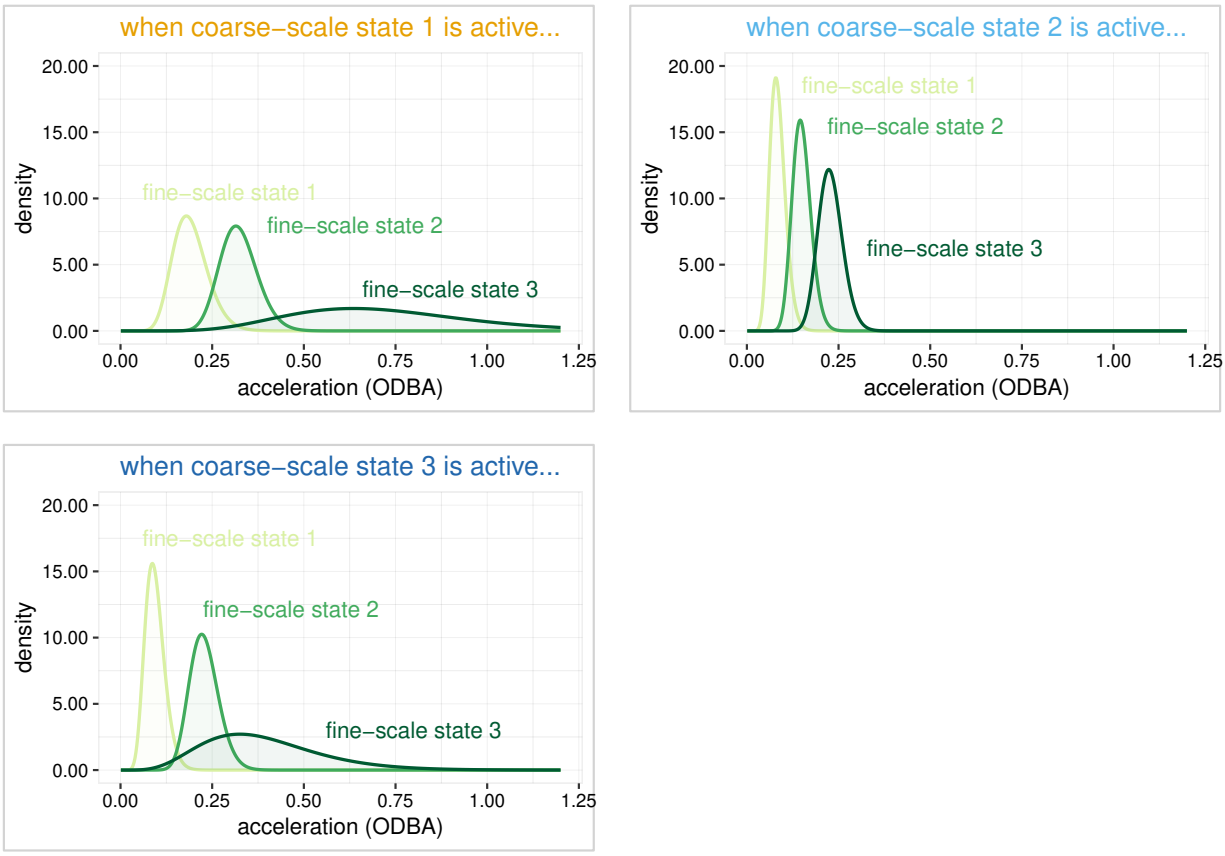

Figure 8: Estimated state-dependent distributions of *fine*-scale acceleration of a horn shark for *coarse*-scale state 1 (zero distance travelled), 2 (zero distance travelled), and 3 (travelling), respectively. *Fine*-scale states 1, 2, and 3 represent *relatively* low, moderate, and high levels of acceleration, respectively, where the corresponding levels substantially differ across *coarse*-scale states (the means of the state-dependent distributions for *fine*-scale state 3, for instance, vary from 0.191 in *coarse*-scale state 2 over 0.323 in *coarse*-scale state 3 to 0.722 in *coarse*-scale state 1).

was frequently in *coarse*-scale state 1 (zero distance travelled, high activity). This area at Santa Catalina Island is composed of smaller reefs in relatively close proximity. Horn sharks are abundant in this area with a relatively high, annual site-fidelity (Strong, 1989), therefore understanding their level of activity, especially in regards to foraging behaviours, can lead to better predictions on how rocky reef communities will be affected.

Quantile-quantile plots and autocorrelation functions of normal pseudo-residuals for three example sequences of *fine*-scale accelerations, each computed as described in Section

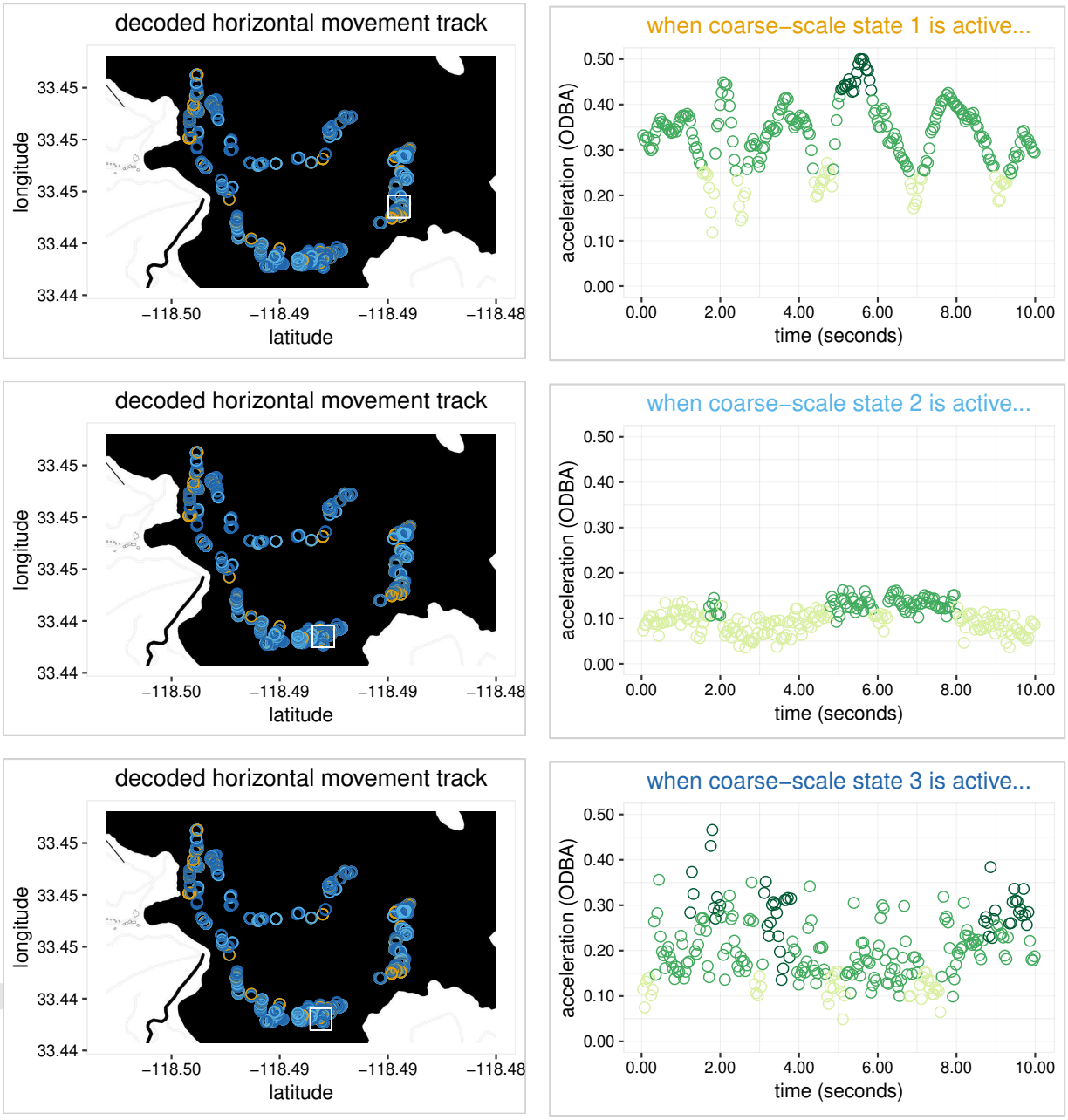

Figure 9: Decoded *coarse*-scale horizontal movement track (left panel) and three example sequences of decoded *fine*-scale accelerations (right panel) of a horn shark, one for *coarse*-scale state 1 (zero distance travelled; $t = 462, \ldots, 466$), one for *coarse*-scale state 2 (zero distance travelled; $t = 1202, \ldots, 1206$), and one for *coarse*-scale state 3 (travelling; $t = 1220, \ldots, 1224$), each of length 10 seconds. The white squares on the maps in the left panel indicate the time intervals which were chosen for the three example sequences displayed in the right panel.

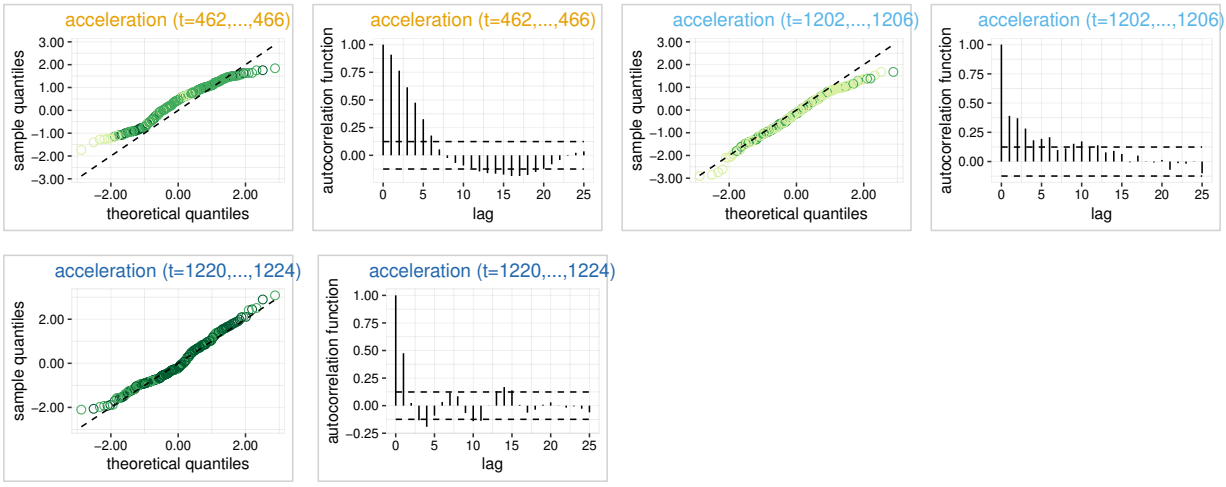

Figure 10: Quantile-quantile plots and autocorrelation functions of normal pseudo-residuals for three example sequences of *fine*-scale accelerations of a horn shark, one for *coarse*-scale state 1 (zero distance travelled; $t = 462, \ldots, 466$), one for *coarse*-scale state 2 (zero distance travelled; $t = 1202, \ldots, 1206$), and one for *coarse*-scale state 3 (travelling; $t = 1220, \ldots, 1224$), each of length 10 seconds. The three example sequences considered here are the same as those displayed in Figure 9, where the colours indicate the Viterbi-decoded states.

2.6, are displayed in Figure 10. (The high proportion of zero step lengths within the *coarse*-scale process renders it difficult to assess the corresponding goodness of fit using pseudo-residuals, such that we show residual plots only for the *fine*-scale data.) The autocorrelation function of the pseudo-residuals obtained for the *fine*-scale process at $t = 462, \ldots, 466$ indicates a violation of the conditional independence assumption across observations. While the HHMM is able to distinguish between active and non-active behaviour when the estimated change in geoposition is zero, the model lacks some structure required for the patterns observed in the acceleration data. However, the quantile-quantile plots for $t = 1202, \ldots, 1206$ and $t = 1220, \ldots, 1224$ show a much better model fit. For ease of model presentation, we did not incorporate additional structure into the HHMM and leave it as an area in which we can continue to develop this modelling framework for application to acceleration data.

4 Discussion

Tracking individual animals can help us to develop a better understanding of an animal's behavioural decisions and the internal and external drivers thereof, and hence in principle also to infer properties of a species. A key aspect in any such analysis is the temporal resolution at which observations are made. A coarse resolution, e.g. with daily observations, can be suitable when the focus lies on migratory patterns, whereas fine-scale data, e.g. as often collected with accelerometers, can reveal detailed information up to individual foraging attempts. While high-resolution data is in principle more informative, here we argue that some of the corresponding short-term decisions made by an animal have to be seen relative to the current context. For example, as shown in the analysis of the Atlantic cod track, complex *fine*-scale movement patterns, such as the effect of the time of day on vertical movement, would not have been revealed without taking the general *coarse*-scale behavioural context (resting/foraging vs. travelling/migrating) into consideration. *Vice versa*, to obtain a more detailed understanding of movement patterns that appear to manifest themselves at coarser scales, it will often be helpful to be able to additionally "zoom in" at a much finer scale. For example, for marine animals, conventional approaches to movement modelling using positional data alone typically aim to infer a combination of resting, area-restricted search, and travelling behaviours. However, as demonstrated in the analysis of the horn shark track, resting, as exemplified by no change in geoposition, is not a valid interpretation for animals that are demersal, non-obligate ram ventilators. In particular, we show that horn sharks make frequent changes across the travelling and not travelling states, and further may be quite active even when they are not actively travelling. In this case, using both *fine*-scale data, such as acceleration signals to inform activity, but also positional data at a relatively coarser time scale to infer whether the shark was actively travelling or not, was necessary to more fully understand the horn shark behaviour.

Fortunately, new types of remote sensing data, in particular such that result from outfitting individual animals with multiple sensors — potentially sampling at differing temporal resolutions — in principle do give us the opportunity to draw much more comprehensive pictures of the dynamics of an animal's behavioural modes. However, these new types of data are rather challenging from the statistical point of view. Our work provides a new modelling framework, the class of HHMMs, which due to its intuitive appeal, its versatility in accommodating various dependence structures and essentially any type of time series data, and the relative ease with which it can be implemented, in principle seems well-suited to handle such data and allow for comprehensive ecological inference from multi-stream and multi-scale data. Unlike previous approaches based on state-space models (SSMs; Jonsen *et al.*, 2005; Auger-Méthé *et al.*, 2016), in the HHMM formulations considered in our case studies we do not explicitly account for measurement error, which can in fact be large in particular for the Atlantic cod location data, which is itself the output of a geolocation model. Depending on the magnitude of the error, failing to propagate this uncertainty through to the statistical model for movement and behaviour can affect state predictions, and hence ultimately also biological inference. Despite this caveat, we see strong potential for HHMMs to become increasingly important in the future, especially due to the ongoing technological progress in biologging technology.

However, the immense flexibility of HHMMs may sometimes also be a curse: challenges which are inherent in basic HMMs, e.g. model selection regarding the number of states or the dependence structure assumed (Pohle *et al.*, 2017), are inevitably exacerbated by the more complex structure of HHMMs. The likelihood-based inferential framework developed in this paper does in principle lend itself to straightforward application of information criteria. However, we strongly advise against over-reliance on such criteria, and instead recommend a more pragmatic approach to finding a suitable model, where expert knowledge of ecologists, thorough exploratory data analysis, and close inspection of how well any plausible candidate model captures the key patterns of interest and relevance, to-

gether guide (and justify) the model selection process. Similarly, while formal approaches such as pseudo-residual analyses to model checking in (H)HMMs do exist, it will not usually be feasible to make a simple binary decision on whether or not a model is suitable based on these, since i) for data as complex such as those considered in our case studies, any reasonably simple statistical model will be deemed inadequate, and ii) unlike in basic HMMs, the model checking in HHMMs applies to different layers, which complicates the decision on the model's suitability. Again, we advocate a holistic approach to checking a fitted model, which most importantly must not lose sight of the actual research question of interest. The "curse of complexity" may have consequences also on the interpretation of an HHMM. While for basic HMMs, when applied to animal movement data, it is often straightforward to at least approximately link model states to biologically meaningful behaviours — e.g. resting, foraging, and travelling when fitting a 3-state HMM to movement data of a terrestrial animal — this may be more problematic within HHMMs, where interpretations ought to be made at different scales.

Regarding the search for an adequate model, we do in fact observe a trend in the field of movement ecology where well-established practices of statistical modelling — exploratory data analysis, formulation of candidate models, selection between candidate models based on some criterion, and checking of the model selected — are increasingly often replaced by rather ad-hoc choices on the (often very complex) model used for inference. Clearly, this development is a reason to be concerned. On the other hand, given the fast-increasing complexity of telemetry data, it comes as no surprise that the development of flexible yet reliable and robust statistical modelling routines are somewhat lacking behind (cf. McClintock *et al.*, 2012). Thus, what we argue for here is to first of all acknowledge the great challenge of finding a suitable statistical model for any given animal telemetry data, and then as a consequence be very thorough and transparent about modelling decisions made, irrespective of whether their justification is based on formal criteria or on more subjective assessments.

The model formulation presented in this paper provides a basic framework for modelling behavioural state switches underlying data collected at multiple scales. A simplifying assumption that was implicitly made within our case studies was that the measurement error was negligible relative to the scales at which movements were observed. This assumption can in principle be relaxed in much the same way as is commonly done in SSMs, by including the actual locations — or, more generally, metrics of interest — in the hidden components of the HHMM, with the state-dependent processes then describing the measurement error. The main disadvantage of such an approach would be the increased computational complexity (Patterson *et al.*, 2017). The hierarchical structure of the model could also be modified with regard to the number of layers considered. Specifically, in the present work we focused on the case of two differing temporal resolutions, leading to two hierarchically structured state processes. However, in analogy to speech recognition, further layers could be introduced; e.g., there could be three layers, corresponding to presence/absence of migratory behaviour (*coarsest* scale), foraging, resting, and travelling modes (*medium* scale), and movements of individual body parts (*finest* scale). It seems intriguing to potentially be able to fit such "complete" models of animal movement, on the other hand such models would certainly not be straightforward to implement and to handle. In such an extension, but also for the models presented in the current paper, an important question is that of optimal statistical design. Specifically, it is of interest to provide general recommendations as to which temporal resolution is needed at either scale in order to answer the research question at hand. The issue of adequately designing animal movement studies is in fact underappreciated in biologging in general (see Patterson *et al.*, 2017).

Acknowledgements

TA and VLB were funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) — Projektnummer 316099922 — TRR 212; CAG was funded by a NERC CASE studentship via the ACCE doctoral training partnership (grant no. NE/L00 2450/1); ENM and CGL received funding from the CSULB Shark Lab. The authors wish to thank Lennart Oelschläger for suggesting valuable improvements with regard to the description of the forward algorithm provided in the Appendix.

Author's contribution

TA and VLB conceived the ideas and designed the methodology; ENM, CGL, and DR collected the data; TA, CAG, VLB, and ENM analysed the data; TA and RL led the writing of the manuscript; PGB contributed major improvements to the manuscript. All authors contributed critically to the drafts and gave final approval for publication.

Data accessibility

The data considered in the real-data applications presented in Section 3 are publicly accessible: the Atlantic cod data is archived on the CEFAS data hub (https://doi.org/10.14466/CefasDataHub.71); the horn shark data is archived on DRYAD (https://doi.org/10.5061/dryad.j86198f).

References

Adlerstein, S.A. & Welleman, H.C. (2000): Diel variation of stomach contents of North Sea cod (*Gadus morhua*) during a 24-h fishing survey: an analysis using generalized additive models. *Canadian Journal of Fisheries and Aquatic Sciences*, **57** (12), 2363–2367.

Auger-Méthé, M., Field, C., Albertsen, C.M., Derocher, A.E., Lewis, M.A., Jonsen, I.D. & Flemming, J.M. (2016): State-space models' dirty little secrets: even simple linear Gaussian models can have estimation problems. *Scientific Reports*, **6**, 26677.

Compagno, L.J.V. (2001): Sharks of the world. An annotated and illustrated catalogue of shark species known to date. Bullhead, mackerel and carpet sharks (*Heterodontiformes*, *Lamniformes* and *Orectolobiformes*). *FAO Species Catalogue for Fishery Purposes*, **1** (2).

DeRuiter, S.L., Langrock, R., Skirbutas, T., Goldbogen, J.A., Calambokidis, J., Friedlaender, A.S. & Southall, B.L. (2016): A multivariate mixed hidden Markov model for blue whale behaviour and responses to sound exposure. *Annals of Applied Statistics*, **11** (1), 362–392.

Fine, S., Singer, Y. & Tishby N. (1998): The hierarchical hidden Markov model: Analysis and applications. *Machine Learning*, **32** (1), 41–62.

Grecian, W.J., Lane, J.V., Michelot, T., Wade, H.M. & Hamer, K.C. (2018): Understanding the ontogeny of foraging behaviour: insights from combining marine predator bio-logging with satellite-derived oceanography in hidden Markov models. *Journal of the Royal Society Interface*, **15** (143), 1–9.

Griffiths, C.A., Patterson, T.A., Blanchard, J.L., Righton, D., Wright, S.R., Pitchford, J.W. & Blackwell, P.G. (2018) Scaling marine fish movement behavior from individuals to populations. *Ecology and Evolution*, **8** (14), 7031—7043.

Hays, C.G., Bailey, H., Bograd, S.J., Don Bowen, W., Campagna, C., Carmichael, R.H., Casale, P., Chiaradia, A., Costa, D.P., Cuevas, E., de Bruyn, P.J.N., Dias, M.P., Duarte, C.M., Dunn, D.C., Dutton, P.H., Esteban, N., Friedlaender, A., Goetz, K.T., Godley, B.J., Halpin, P.N., Hamann, M., Hammerschlag, N., Harcourt, R., Harrison, A.-L., Hazen, E.L., Heupel, M.R., Hoyt, E., Humphries, N.E., Kot, C.Y., Lea, J.S.E., Marsh,

H., Maxwell, S.M., McMahon, C.R., Notarbartolo di Sciara, G., Palacios, D.M., Phillips, R.A., Righton, D., Schofield, G., Seminoff, J.A., Simpfendorfer, C.A., Sims, D.W., Takahashi, A., Tetley, M.J., Thums, M., Trathan, P.N., Villegas-Amtmann, S., Wells, R.S., Whiting, S.D., Wildermann, N.E. & Sequeira, A.M.M. (2019): Translating marine animal tracking data into conservation policy and management. *Trends in Ecology & Evolution*, **2491**, 1–15.

Hobson, V.J., Righton, D., Metcalfe, J.D. & Hays, G.C. (2007): Vertical movements of North Sea cod. *Marine Ecology Progress Series*, **347**, 101–110.

Hobson, V.J., Righton, D., Metcalfe, J.D. & Hays, G.C. (2009): Link between vertical and horizontal movement patterns of cod in the North Sea. *Aquatic Biology*, **5**, 133–142.

Hunter, E., Metcalfe, J.D., O'Brien, C.M., Arnold, G.P. & Reynolds, J.D. (2004): Vertical activity patterns of free-swimming adult plaice in the southern North Sea. *Marine Ecology Progress Series*, **279**, 261–273.

Hunter, E., Metcalfe, J.D., Holford, B.H. & Arnold, G.P. (2004): Geolocation of free-ranging fish on the European continental shelf as determined from environmental variables II. Reconstruction of plaice ground tracks. *Marine Biology*, **144**, 787–798.

Hunter, E., Buckley, A.A., Stewart, C. & Metcalfe, J.D. (2005): Migratory behaviour of the thornback ray, *Raja clavata*, in the southern North Sea. *Journal of the Marine Biological Association of the United Kingdom*, **85**, 1095–1105.

Hussey, N.E., Kessel, S.T., Aarestrup, K., Cooke, S.J., Cowley, P.D., Fisk, A.T., Harcourt, R.G., Holland, K.N., Iverson, S.J., Kocik, J.F., Mills Flemming, J.E. & Whoriskey, F.G. (2015): Aquatic animal telemetry: a panoramic window into the underwater world. *Science*, **348** (6240), 1255642.

ICES (2018): *Report of the Working Group on Assessment of Demersal Stocks in the North Sea and Skagerrak (WGNSSK)*. ICES CM 2018/ACOM:22.

Jonsen, I., Flemming, J. & Myers, R. (2005): Robust state-space modeling of animal movement data. *Ecology*, **86** (11), 2874–2880.

Johnson, D., London, J., Lea, M. & Durban, J. (2008): Continuous-time correlated random walk model for animal telemetry data. *Ecology*, **89** (5), 1208–1215.

Langrock, R., Marques, T.A., Baird, R.W. & Thomas, L. (2013): Modeling the diving behavior of whales: a latent-variable approach with feedback and semi-Markovian components. *Journal of Agricultural, Biological and Environmental Statistics*, **19** (1), 82–100.

Langrock, R., Adam, T., Leos-Barajas, V., Mews, S., Miller, D.L. & Papastamatiou, Y.P. (2018): Spline-based nonparametric inference in general state-switching models. *Statistica Neerlandica*, **72** (3), 179–200.

Leos-Barajas, V., Photopoulou, T., Langrock, R., Patterson, T.A., Watanabe, Y.Y., Murgatroyd, M. & Papastamatiou, Y.P. (2017a): Analysis of animal accelerometer data using hidden Markov models. *Methods in Ecology and Evolution*, **8** (2), 161–173.

Leos-Barajas, V., Gangloff, E.J., Adam, T., Langrock, R., van Beest, F.M., Nabe-Nielsen, J. & Morales, J.M. (2017b): Multi-scale modeling of animal movement and general behavior data using hidden Markov models with hierarchical structures. *Journal of Agricultural, Biological and Environmental Statistics*, **22** (3), 232–248.

Løkkeborg, S. (1998): Feeding behaviour of cod, *Gadus morhua*: activity rhythm and chemically mediated food search. *Animal Behaviour*, **56** (2), 371–378.

Love, M.S. (1991): *Probably more than you want to know about the fishes of the Pacific coast, 2nd edition*. Really Big Press, Santa Barbara.

Li, M. & Bolker, B.M. (2017): Incorporating periodic variability in hidden Markov models for animal movement. *Movement Ecology*, **5**:1.

McClintock, B.T., King, R., Thomas, L., Matthiopoulos, J., McConnell, B.J. & Morales, J.M. (2012): A general discrete-time modeling framework for animal movement using multistate random walks. *Ecological Monographs*, **82** (3), 335–349.

Michelot, T., Langrock, R. & Patterson, T.A. (2016): moveHMM: an R package for analysing animal movement data using hidden Markov models. *Methods in Ecology and Evolution*, **7** (11), 1308–1315.

Michelot, T., Langrock, R., Bestley, S., Jonsen, I.D., Photopoulou, T. & Patterson, T.A. (2017): Estimation and simulation of foraging trips in land-based marine predators. *Ecology*, **98** (7), 1932–1944.

Morales, J.M., Haydon, D.T., Frair, J., Holsinger, K.E. & Fryxell, J.M. (2004): Extracting more out of relocation data: building movement models as mixtures of random walks. *Ecology*, **85** (9), 2436-–2445.

Neat, F.C., Bendall, V., Berx, B., Wright, P.J., Cuaig, M., Townhill, B., Schön, P.-J., Lee, J. & Righton, D. (2014): Movement of Atlantic cod around the British Isles: implications for finer scale stock management. *Journal of Applied Ecology*, **51** (6), 1564-–1574.

Nelson, D. & Johnson, R. (1970): Diel activity rhythms in the nocturnal bottom-dwelling sharks, *Heterodontus francisci* and *Cephaloscyllium ventriosum*. *Copeia*, **1970** (4), 732-–739.

Patterson, T.A., Basson, M., Bravington, M.V. & Gunn, J.S. (2009): Classifying movement behaviour in relation to environmental conditions using hidden Markov models. *Journal of Animal Ecology*, **78** (6), 1113–1123.

Patterson, T.A., Parton, A., Langrock, R., Blackwell, P.G., Thomas, L. & King, R. (2017): Statistical modelling of individual animal movement: an overview of key methods and a discussion of practical challenges. *AStA Advances in Statistical Analysis*, **101** (4), 399–438.

Pedersen, M.W., Righton, D., Thygesen, U.H., Andersen, K.H. & Madsen, H. (2008): Geolocation of North Sea cod (*Gadus morhua*) using hidden Markov models and behavioural switching. *Canadian Journal of Fisheries and Aquatic Sciences*, **65** (11), 2367–2377.

Pohle, J., Langrock, R., van Beest, F.M. & Schmidt, N.M. (2017): Selecting the number of states in hidden Markov models – pragmatic solutions illustrated using animal movement. *Journal of Agricultural, Biological and Environmental Statistics*, **22** (3), 270–293.

R Core Team (2017): R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. https://www.r-project.org/.

Righton, D., Metcalfe, J.D. & Connolly, P. (2001): Different behaviour of North and Irish Sea cod. *Nature*, **411** (6834), 156.

Righton, D., Quayle, V.A., Hetherington, S. & Burt, G. (2007): Movements and distribution of cod (*Gadus morhua*) in the southern North Sea and English Channel: results from conventional and electronic tagging experiments. *Journal of the Marine Biological Association of the United Kingdom*, **87** (2), 599–613.

Rutz, C. & Hays, G.C. (2009): New frontiers in biologging science. *Biology Letters*, **5**, 289–292.

Shepard, E.L.C., Wilson, R., Quintana, F., Laich, A.G., Liebsch, N., Albareda, D.A., Halsey, L.G., Gleiss, A.C., Morgan, D.T., Myers, A.E., Newman, C. & McDonald,

D.C. (2008): Identification of animal movement patterns using tri-axial accelerometry. *Endangered Species Research*, **10**, 47–60.

Strong, W. (1989): *Behavioral ecology of horn sharks, Heterodontus francisci, at Santa Catalina Island, California, with emphasis on patterns of space utilization*. Department of Biological Sciences, California State University, Long Beach.

Whoriskey, K., Auger-Méthé, M., Albertsen, C.M., Whoriskey, F.G., Binder, T.R., Krueger, C.C. & Mills Flemming, J. (2017): A hidden Markov movement model for rapidly identifying behavioral states from animal tracks. *Ecology and Evolution*, **7** (7), 2112–2121.

Wilson, R.P., White, C.R., Quintana, F., Halsey, L.G., Liebsch, N., Martin, G.R. & Butler, P.J. (2006): Moving towards acceleration for estimates of activity-specific metabolic rate in free-living animals: the case of the cormorant. *Journal of Animal Ecology*, **75** (5), 1081–1090.

Zucchini, W., MacDonald, I.L. & Langrock, R. (2016): *Hidden Markov models for time series: an introduction using R, 2nd edition*. Chapman & Hall/CRC, Boca Raton.

Appendix

To evaluate the logarithm of the likelihood given in (5), we proceed as follows: first, we evaluate the log-likelihoods of the *fine*-scale observations, corresponding to each of the T chunks of *fine*-scale observations being generated by each of the N *fine*-scale HMMs as selected by the *coarse*-scale state process.

Let $\phi^{*(i)}_{l^*,t^*} = \log\big(f^{(i)}(\mathbf{y}^*_{t,1},\ldots,\mathbf{y}^*_{t,t^*}, s^*_{t,t^*}=l^*)\big)$, $l^*=1,\ldots,N^*$, denote the so-called *fine*-scale log-forward probabilities under the i-th *fine*-scale HMM. The *fine*-scale log-forward probabilities can be calculated via the forward algorithm, which amounts to applying the recursion

$$\begin{aligned}
\phi^{*(i)}_{l^*,1} &= \log\left(\delta^{*(i)}_{l^*} f^{*(i,l^*)}(\mathbf{y}^*_{t,1})\right) \\
&= \log\left(\delta^{*(i)}_{l^*}\right) + \log\left(f^{*(i,l^*)}(\mathbf{y}^*_{t,1})\right) \\
\phi^{*(i)}_{l^*,t^*} &= \log\left(\sum_{k^*=1}^{N^*} \exp\left(\phi^{*(i)}_{k^*,t^*-1}\right)\gamma^{*(i)}_{k^*l^*} f^{*(i,l^*)}(\mathbf{y}^*_{t,t^*})\right) \\
&= \log\left(\sum_{k^*=1}^{N^*} \exp\left(\phi^{*(i)}_{k^*,t^*-1} + \log\left(\gamma^{*(i)}_{k^*l^*}\right) - c_{t-1}\right)\right) + c_{t-1} + \log\left(f^{*(i,l^*)}(\mathbf{y}^*_{t,t^*})\right),
\end{aligned}$$

$t^*=2,\ldots,T^*$, where $c_t = \max(\phi^{*(i)}_{1,t^*},\ldots,\phi^{*(i)}_{N^*,t^*})$ is a constant that is used within the log-sum-of-exponentials-function to prevent numerical underflow which may occur when exponentiating large negative numbers.

Since $\mathcal{L}^{\mathrm{HMM}}_i(\mathbf{y}^*_t) = f^{(i)}(\mathbf{y}^*_t) = \sum_{l^*=1}^{N^*} f^{(i)}(\mathbf{y}^*_t, s^*_{t,T^*}=l^*)$, the log-likelihood of the t-th chunk of fine-scale observations being generated by the i-th fine-scale HMM follows as

$$\log\left(\mathcal{L}^{\mathrm{HMM}}_i(\mathbf{y}^*_t)\right) = \log\left(\sum_{l^*=1}^{N^*} \exp\left(\phi^{*(i)}_{l^*,T^*} - c_T\right)\right) + c_T. \tag{1}$$

After having evaluated the log-likelihoods (1) for each of the T chunks of *fine*-scale observations and N *fine*-scale HMMs, we calculate the so-called *coarse*-scale log-forward

probabilities $\phi_{j,t} = \log\big(f(\mathbf{y}_1^*, \ldots, \mathbf{y}_t^*, \mathbf{y}_1, \ldots, \mathbf{y}_t, s_t = j)\big)$, $j = 1, \ldots, N$, which can be calculated in a similar way as in () by applying the recursion

$$\begin{aligned}
\phi_{j,1} &= \log\left(\delta_j \mathcal{L}_j^{\mathrm{HMM}}(\mathbf{y}_1^*) f^{(j)}(\mathbf{y}_1)\right) \\
&= \log(\delta_j) + \log\left(\mathcal{L}_j^{\mathrm{HMM}}(\mathbf{y}_1^*)\right) + \log\left(f^{(j)}(\mathbf{y}_1)\right) \\
\phi_{j,t} &= \log\left(\sum_{i=1}^{N} \exp\left(\phi_{i,t-1}\right) \gamma_{ij} \mathcal{L}_j^{\mathrm{HMM}}(\mathbf{y}_t^*) f^{(j)}(\mathbf{y}_t)\right) \\
&= \log\left(\sum_{i=1}^{N} \exp\left(\phi_{i,t-1} + \log(\gamma_{ij}) - c_{t-1}\right)\right) + c_{t-1} + \log\left(\mathcal{L}_j^{\mathrm{HMM}}(\mathbf{y}_t^*)\right) + \log\left(f^{(j)}(\mathbf{y}_t)\right),
\end{aligned}$$

$t = 2, \ldots, T$, where $c_t = \max(\phi_{1,t}, \ldots, \phi_{N,t})$.

Since $\mathcal{L}^{\mathrm{HHMM}}(\mathbf{y}, \mathbf{y}^*) = f(\mathbf{y}, \mathbf{y}^*) = \sum_{j=1}^{N} f(\mathbf{y}, \mathbf{y}^*, s_T = j)$, the log-likelihood of the HHMM follows as

$$\log\left(\mathcal{L}^{\mathrm{HHMM}}(\mathbf{y}, \mathbf{y}^*)\right) = \log\left(\sum_{j=1}^{N} \exp(\phi_{j,T} - c_T)\right) + c_T.$$