

Hidden Markov Models: Pitfalls and Opportunities in Ecology

Richard Glennie¹, Timo Adam¹, Vianey Leos-Barajas², Théo Michelot¹, Theoni Photopoulou¹, and Brett T. McClintock³

¹Centre for Research into Ecological and Environmental Modelling, University of St Andrews, St Andrews, KY16 9LZ, UK

²Department of Statistical Sciences, University of Toronto, Toronto, ON, Canada

³Marine Mammal Laboratory, NOAA-NMFS Alaska Fisheries Science Center, Seattle, USA

26th December 2021

Abstract

- (1) Hidden Markov models (HMMs) and their extensions are attractive methods for analysing ecological data where noisy, multivariate measurements are made of a hidden, ecological process, and where this hidden process is represented by a sequence of discrete states. Yet, as these models become more complex and challenging to understand, it is important to consider what pitfalls these methods have and what opportunities there are for future research to address these pitfalls.
- (2) In this paper, we review five lesser known pitfalls one can encounter when using HMMs or their extensions to solve ecological problems: (1) violation of the snapshot property in continuous-time HMMs; (2) biased inference from hierarchical HMMs when applied to temporally misaligned processes; (3) sensitive inference from using random effects to partially pool across heterogeneous individuals; (4) computational burden when using HMMs to approximate models with continuous state spaces; and (5) difficulty linking the hidden process to space or environment.
- (3) This review is for ecologists and ecological statisticians familiar with HMMs, but who may be less aware of the problems that arise in more specialised applications. We demonstrate how each pitfall arises, by simulation or example, and discuss why this pitfall is important to consider. Along with

identifying the problems, we highlight potential research opportunities and offer ideas that may help alleviate these pitfalls.

(4) Each of the methods we review are solutions to current ecological research problems. We intend for this paper to heighten awareness of the pitfalls ecologists may encounter when applying these more advanced methods, but we also hope that by highlighting future research opportunities, we can inspire ecological statisticians to weaken these pitfalls and provide improved methods.

Keywords: animal movement, continuous time, hidden Markov model, hierarchical model, population ecology, random effects, state space models, time series

1 Introduction

The statistical methods used in ecology are becoming increasingly complex. What statistical and computational pitfalls do these methods have? What future research opportunities are there for these methods to be improved? In this paper, we discuss these two questions for a popular class of statistical models: hidden Markov models (HMMs; Zucchini et al., 2017).

HMMs are widely applied in ecology, from individual-level to ecosystem-level modelling (see McClintock et al. (2020) for a comprehensive overview of the application of HMMs in ecology). Their intuitive structure often corresponds with our conceptual models for ecological systems: there is a hidden process unfolding over time from which we obtain noisy, multivariate observations. In their commonest form, HMMs consist of two time series (S_t, \mathbf{Y}_t) , in discrete time, where S_t is a hidden (unobserved) state that can take one of a finite number of values and \mathbf{Y}_t is a collection of observed variables whose distribution we assume depends on the hidden state S_t , termed its state-dependent distribution. Crucially, it is assumed the observations are independent given the states. Over time, states evolve as a Markov process which is described by a transition probability matrix Γ_t whose $(i, j)^{\text{th}}$ entry is the probability that $S_{t+1} = j$ given $S_t = i$. Key

44 properties of HMMs are that they commonly deal with observations in regular, discrete time; the state, S_t ,
45 switches between a pre-specified finite number of possible values; and there are computationally efficient
46 algorithms to quickly fit HMMs to large amounts of data.

47 The successful application of HMMs in ecology has inspired further use that goes beyond the standard
48 application of HMMs, and it is these extended uses that this paper will discuss (Readers unfamiliar with
49 HMMs are encouraged to consult McClintock et al. (2020) or Zucchini et al. (2017) for an introduction).
50 We consider five extended uses of HMMs: (1) HMMs in continuous-time; (2) hierarchical HMMs; (3)
51 HMMs with random effects; (4) approximating state space models with HMMs; and (5) two-stage analyses
52 of decoded states from HMMs. Each extended use adds complexity to what is already a complex statistical
53 analysis and so we ask the two questions above: what pitfalls are revealed and what future research
54 opportunities are there? We hope this discussion will serve two audiences. For ecologists familiar with
55 HMMs, it provides a statistical overview of more advanced uses of HMMs, focusing on what possible
56 problems a practitioner may encounter. For statistical ecologists or ecological statisticians, we hope this
57 paper serves as a resource to inspire future research, either to weaken the pitfalls we highlight in HMMs
58 or to introduce alternative methods to solve these problems.

59 **2 Continuous-time hidden Markov models**

60 **2.1 Introduction**

61 A continuous-time HMM has the same dependence structure as a discrete-time HMM, but the underlying
62 state S_t of the system is determined by a continuous-time Markov chain. An N -state continuous-time
63 Markov chain is defined by an $N \times N$ transition *rate* matrix,

$$Q = \begin{pmatrix} -q_1 & q_{12} & \cdots & q_{1N} \\ q_{21} & -q_2 & \cdots & q_{2N} \\ \vdots & \vdots & \ddots & \vdots \\ q_{N1} & q_{N2} & \cdots & -q_N \end{pmatrix}$$

64 where $q_i = \sum_{j \neq i} q_{ij}$ for all i . The time spent in each state i follows an exponential distribution
 65 with mean $1/q_i$. The non-diagonal elements of the matrix are strictly positive and proportional to the
 66 transition probabilities out of state i (within each row). For any given time interval of length $\Delta > 0$, the
 67 corresponding transition probability matrix over that time interval can be computed from the transition
 68 rate matrix with a matrix exponential:

$$\Gamma_{\Delta} = e^{Q\Delta}. \tag{1}$$

69 The key property of continuous-time HMMs is that they do not require observations made at regular time
 70 intervals, and the times of observations do not need to match the times of state transitions. A schematic
 71 representation of the dependence structure of a continuous-time HMM is shown in Fig. 1.

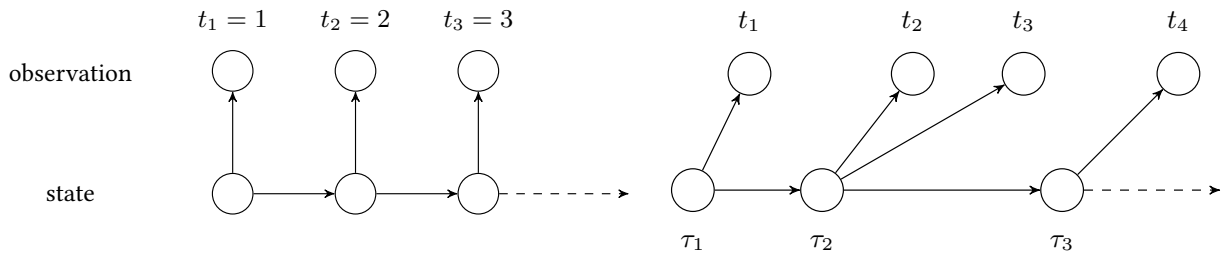


Fig. 1. Comparison of dependence graphs for discrete-time (left) and continuous-time (right) HMMs. In both cases, the distribution of an observation only depends on the current value of the state. The key difference is that, in a continuous-time HMM, the times of state transitions (τ_1, τ_2, \dots) and the times of observation (t_1, t_2, \dots) do not need to match, and both may be irregularly spaced.

72 Continuous-time HMMs have been popular in medical statistics, in particular to study disease progres-
 73 sion in patients from consultations done at irregular time intervals (e.g., Jackson et al., 2003; Bureau et al.,

2003; Liu et al., 2015). However, they remain rare in ecological studies, despite their broad applicability to irregular data sets. We posit that they have been underutilised because the mathematical theory behind continuous-time Markov processes is less intuitive than its counterpart in discrete time, and because of the apparent lack of accessible software to readily apply this method. However, much of the inferential framework developed for discrete-time HMMs can also be used in continuous time, based on the relationship between transition rates and transition probabilities in Equation (1). This includes the forward algorithm and the Viterbi algorithm (Zucchini et al., 2017). In many situations, the implementation of a continuous-time HMM is virtually identical to that of a discrete-time HMM, with the only difference being that the likelihood function is parametrised in terms of transition rates. Yet, it is once you move from discrete to continuous time that a central property of HMMs becomes less intuitive: the snapshot property.

2.2 Pitfall: snapshot property violation

The snapshot property is satisfied if the value of the observation process at time t only depends on the state at that time, rather than on past values of the state (Patterson et al., 2017). In discrete-time, this is a well-understood assumption; however, in continuous-time, state transitions could happen *at any time* (Fig. 1). Continuous-time HMMs are therefore only suitable when the distribution of each observation does not depend on the intermediate sequence of switches that have occurred since the previous observation. When this is not the case, the snapshot property may still be a reasonable approximation if observations occur at a high temporal resolution relative to the scale of state switching, i.e., if only a small proportion of observation intervals contain state switches. We illustrate this below using simulations.

One context where the snapshot property is usually violated is state-switching models based on relatively infrequent observations of animal locations or velocities, because these variables depend on the

96 behaviour of the animal over the whole time interval between observations rather than at the time of
97 observation (Blackwell et al., 2016; Patterson et al., 2017; Michelot and Blackwell, 2019). We therefore
98 investigated the approximation error arising from using a continuous-time HMM in cases where the
99 snapshot property is not satisfied. As a simple example of a movement model, we considered a 2-state
100 Brownian motion process with no drift, where the diffusion parameter was small in state $S_t = 1$ (slow
101 movement) and large in $S_t = 2$ (fast movement). To facilitate interpretation of the simulations, we used
102 parameters estimated from a real data set of Antarctic petrels (*Thalassoica antarctica*) from the Movebank
103 data repository (Descamps et al., 2016a,b). The diffusion parameters estimated from the petrel data were
104 $(\sigma_1, \sigma_2) = (1.2, 14.9)$ (i.e., mean distance travelled over 1 hour was 1.5km in state 1 and 18.7km in state
105 2), and the transition rates were $q_{12} = 0.29$, and $q_{21} = 0.33$ (i.e., expected dwelling time was 3.4h in state
106 1 and 3h in state 2). The details of the petrel analysis and the formulation of the Brownian motion are
107 given in Section S1.1 of the Supplementary Material.

108 We generated data using those parameters and mimicked the real-life scenario where the behavioural
109 switching times are unknown, to evaluate the error caused by violations of the snapshot property. We then
110 assessed the performance of the approximation based on bias between true and estimated parameters, and
111 on the proportion of correctly-estimated states, for different time intervals of observation. The simulation
112 procedure is described in Section S1.2 of the Supplementary Material, and the results are shown in Fig. 2.
113 As expected, the approximation error increases with the mean time interval because more switches can
114 occur between observations (but each observation is assumed to only depend on the state active at that
115 time).

116 The performance of the continuous-time HMM methodology in contexts where the snapshot property
117 does not hold depends on the time scale at which the hidden states occur (i.e., on the transition rates)
118 relative to the time scale of the observations. Our results support an intuitive rule of thumb for determining

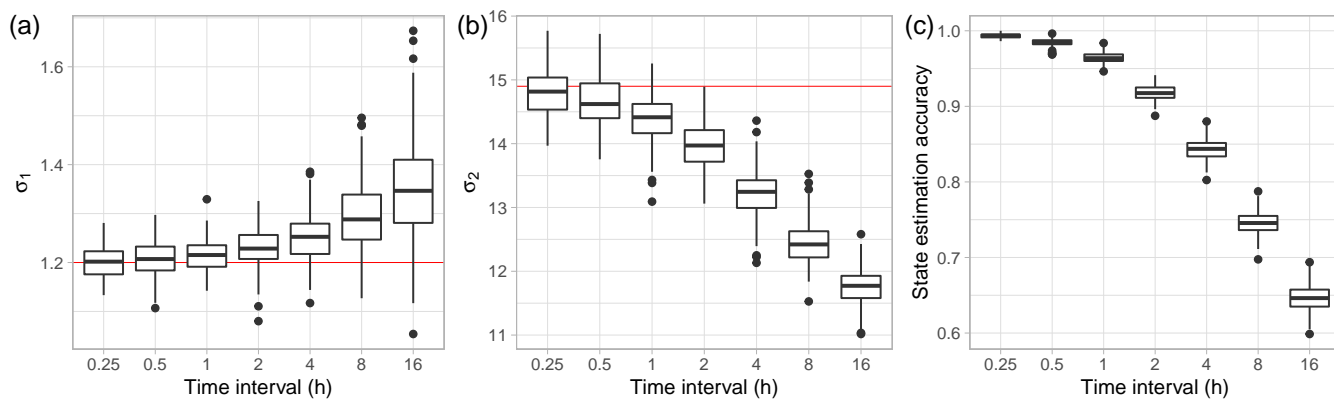


Fig. 2. Results of continuous-time HMM simulation study. Estimated parameters σ_1 (a) and σ_2 (b), and state estimate accuracy (c) for simulated data with different mean time intervals, from 0.25h to 16h. Each box contains 200 replications, for each simulation scenario.

119 whether or not any given sampling interval may reasonably satisfy the snapshot property:

$$\Delta \leq \frac{1}{\max_i (q_i)}, \quad (2)$$

120 i.e., the mean time interval of observation should be shorter than the shortest expected dwelling time
 121 (3h in our example). However, as with all HMMs, performance will also depend on the amount of
 122 state-dependent observation distribution overlap, serial correlation in the hidden state sequence, and
 123 other properties of the data (e.g. Zucchini et al., 2017; McClintock, 2021). These results suggest that
 124 continuous-time HMMs should not be applied naively in ecological studies where the snapshot property
 125 is known to be violated, in particular when the time intervals of observations are long relative to the time
 126 scale of the state process.

127 2.3 Opportunities

128 Our simulation results indicated that it is important to consider whether or not the snapshot property is
 129 reasonably satisfied when using continuous-time HMMs for ecological data. Continuous-time HMMs
 130 have recently been proposed for the analysis of capture-recapture data (Choquet et al., 2017; Mews et al.,
 131 2020a), where the observation is a categorical variable which indicates whether an animal was captured
 132 or not at a given capture occasion, and the hidden state is the existential state of that animal (usually

133 “alive” or “dead”). The probability of capturing the animal at time t only depends on whether the animal is
134 alive at time t , rather than on whether it was alive over the whole interval since the last capture occasion,
135 and the system therefore has the snapshot property. Continuous-time HMMs are a natural framework
136 for this type of data because capture occasions are often irregular in time. Similarly, HMMs developed
137 for occupancy data could be extended to the continuous-time case, because the observation (“detected”
138 or “not detected”) does not depend on the history of the state process, conditional on the current state
139 (“present” or “absent”; see Royle and Kéry, 2007; MacKenzie et al., 2018).

140 For other types of data where the snapshot property cannot be satisfied by study design (e.g., by ensuring
141 time intervals are sufficiently short relative to the transition rates), there are existing remedies. One
142 option is to augment the data with missing observation times (thereby yielding sufficiently short intervals)
143 and then integrate over the missing observations during model fitting. This integration is relatively
144 straightforward in Bayesian analysis (e.g., Gelman et al., 2013) and maximum likelihood analyses using
145 expectation-maximization algorithms (e.g., McLachlan and Krishnan, 2007). When the forward algorithm
146 is used to maximize the likelihood directly, multiple imputation methods can account for the missing
147 observations (e.g. Rubin, 2004; McClintock, 2017). These approaches are approximate but relatively easy
148 to implement. The challenge with exact methods is that the state transition times are unknown. An
149 exact but relatively difficult method to implement is the integrated continuous-time HMM (Blackwell,
150 2018), which is a Bayesian approach that utilizes the efficient forward algorithm to estimate the whole
151 state process together with the model parameters. Further development of exact methods that efficiently
152 account for violations of the snapshot property is a promising area of future research.

153 While discrete-time HMMs for ecological data have been widely applied and extended in recent decades
154 (e.g. McClintock et al., 2020), continuous-time HMMs have received far less attention. This is unfortunate
155 because many types of ecological data are collected in continuous time, but are then “shoe-horned” into a

156 discrete-time framework for analysis (e.g. Borchers et al., 2014; McClintock et al., 2014). Although they
157 have not yet seen wide use by ecologists, there are several R (R Core Team, 2020) packages that make
158 continuous-time HMMs readily applicable to ecological data sets that satisfy the snapshot property. In
159 particular, the package `msm` includes many common observation distributions and allows for covariate
160 dependence in the model parameters (Jackson, 2011). More recently, the package `momentuHMM` has
161 been extended to accommodate continuous-time HMMs, with an emphasis on observation distributions
162 commonly used in animal movement behaviour models for biotelemetry data (McClintock and Michelot,
163 2018). Other options include `HMMCont`, which is limited to normally-distributed observations (Beketov,
164 2014), and JAGS using the `msm` module (Plummer, 2017). In addition to refinements that can more
165 efficiently account for violations of the snapshot property, there remain many opportunities to develop
166 continuous-time analogues to recent extensions of discrete-time HMMs, including semi-Markov models
167 (e.g. Langrock and Zucchini, 2011), hierarchical HMMs (e.g. Fine et al., 1998, see Section 3), random effects
168 (e.g. Altman, 2007, see Section 4), and covariates that vary in continuous time (e.g. Mews et al., 2020a).
169 Such developments will help improve continuous-time HMMs for ecology and facilitate their application
170 to a broad range of systems and taxa.

171 **3 Hierarchical hidden Markov models**

172 **3.1 Introduction**

173 Hierarchical HMMs (HHMMs; Fine et al., 1998) extend basic HMMs by having multiple processes operate
174 at different time scales (Leos-Barajas et al., 2017; Adam et al., 2019). The key property of HHMMs is that
175 they model multi-scale data jointly (e.g. hourly step lengths from GPS tags and accelerations recorded
176 from accelerometers several times per second), that are driven by multiple hidden Markov chains that
177 evolve at different time scales (e.g. behavioural state each hour and within that behavioural substate each

178 second). They are already successfully applied in ecology, e.g., to the dive behaviour of harbour porpoises
179 (Leos-Barajas et al., 2017; Sacchi and Swallow, 2021), to horizontal and vertical movements of Atlantic
180 cod (Adam et al., 2019) and white sharks (Aquino-Baleytó et al., 2021), and to the kinematic movements
181 of northern resident killer whales (Sidrow et al., 2021).

182 In their simplest form, HHMMs have two hierarchical hidden processes, each of which operates at
183 a different time scale: (1) a coarse-scale (e.g. hours) hidden process S_t for coarse-scale time t and (2) a
184 fine-scale (e.g. seconds) hidden process S_{t,t^*} for fine-scale time t^* within coarse-scale time t . Each of these
185 hidden processes has its own transition probability matrix and each can have an associated observation
186 process for observations that occur at either coarse-scale or fine-scale resolution. Essentially, if the
187 coarse-scale state process has N states, the fine-scale observations are modelled by N different fine-scale
188 HMMs. The coarse-scale state that is active determines which fine-scale HMM models the fine-scale
189 observations. As the fine-scale observations do not only depend on the fine-scale state process but also
190 on the coarse-scale state process, HHMMs can capture how switches between fine-scale behaviours
191 depend on the coarse-scale behavioural mode. This added flexibility can then better capture how animals
192 determine their behaviour in both the short and long term (Adam et al., 2019).

193 **3.2 Pitfalls**

194 Not all multi-scale data follow the dependence structure assumed by HHMMs. As described in Section 3.1,
195 a central assumption is that the coarse-scale observations and the fine-scale observations are ultimately
196 driven by a coarse-scale state process; the dependence structure is thus determined by the resolution
197 of the coarse-scale observations (see Fig. 3 (a) for an illustration of two state processes that satisfy this
198 assumption). Consider a scenario with coarse-scale step lengths observed once per hour and fine-scale
199 accelerations observed once per second. HHMMs assume that given the coarse-scale state for that hour,

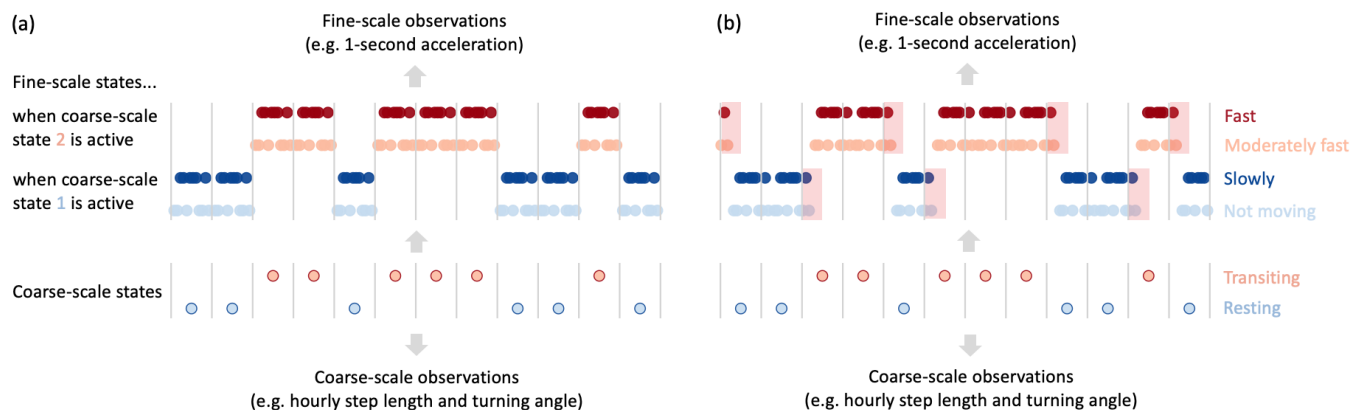


Fig. 3. Illustration of the state processes of an HHMM where the processes are aligned (a) and where this assumption is violated (b). A typical example for the latter scenario is an animal that performs (moderately) fast movements during the first few minutes of an hour that is, overall, characterised by resting behaviour, or no (slow) movements during an hour that is, overall, characterised by transiting behaviour (red-shaded areas). Hourly segments are indicated by vertical lines.

200 the accelerations during that *entire* hour arise from a single fine-scale HMM; in the next hour, the coarse-
 201 scale state may change and so the accelerations may arise from a different HMM. The important insight
 202 is that the accelerations cannot arise from two different fine-scale HMMs within the same hour.

203 While this assumption is reasonable in many scenarios, it may be questionable in other applications.
 204 An animal can perform (moderately) fast movements during the first few minutes of an hour that is,
 205 overall, characterised by resting behaviour, or no (slow) movements during the first few minutes of an
 206 hour that is, overall, characterised by transiting behaviour (see the red-shaded areas in Fig. 3 (b) for an
 207 illustration of such a scenario). Similarly, the fine-scale observations are not necessarily driven by the
 208 *same* coarse-scale state process as the coarse-scale observations. Accelerations can be driven by other
 209 behaviours than resting and transiting, which drive step lengths and turning angles. In both examples,
 210 multi-scale data do not follow the dependence structure that is assumed by HHMMs.

211 To show the consequences of such a violation of the dependence structure, we conducted a simulation
 212 experiment (further details on the simulation procedure are provided in Section S2.1 of the Supplementary
 213 Material). Over 200 replications, we simulated a 2-state coarse-scale process on the hourly scale with

214 1000 observations of step length and turning angle, and a 2-state fine-scale process with 100 observations
215 of acceleration. We then progressively shifted the fine-scale process by 0, 5, 10, 15, and 20 observations
216 and computed the percentage bias in parameter estimates (full descriptions of the parameters are provided
217 in Section S2.1 of the Supplementary Material). While we used deterministic shifts, in practice, they can
218 also be probabilistically. In that regard, the deterministic shifts used for the simulation experiment can be
219 thought of as being exemplary of a probabilistic shifting process with mean equal to the deterministic
220 shifts and small variance. For probabilistic shifting processes with large variance, or scenarios where the
221 fine-scale HMMs change probabilistically within a coarse-scale state, we expect that HHMMs often fail to
222 infer distinct behavioural modes at the fine scale, as all fine-scale behaviours occur within all coarse-scale
223 states, i.e., the two processes are less correlated. In such cases, separate HMMs for the two processes
224 should be preferred over HHMMs. All models were fit using the R package `momentuHMM` (McClintock
225 and Michelot, 2018).

226 Example results are displayed in Fig. 4 (full results are displayed in Section S2.2 of the Supplementary
227 Material). Clearly, when the observations were simulated as assumed by the HHMM (i.e., without shifting
228 the fine-scale process), the estimates are unbiased; however, the question is how severe the bias increases
229 as the fine-scale state process is shifted. While the bias remains relatively small for small shifts, it increases
230 sharply with shifting of the fine-scale process. This severe bias is due to the fact that each of the two
231 fine-scale HMMs must accommodate observations within each hour that truly belong to the alternate
232 fine-scale HMM: a restriction imposed by having an hourly coarse-scale process.

233 As a consequence, ecological interpretations of the state-dependent distributions can be difficult to
234 elicit (in the above example, for instance, the true within-state variance of the accelerations is much
235 smaller than the estimated one, simply because of the temporal regularity of the observations and the
236 dependence structure of the HHMM). Furthermore, pitfalls regarding model selection, which are often

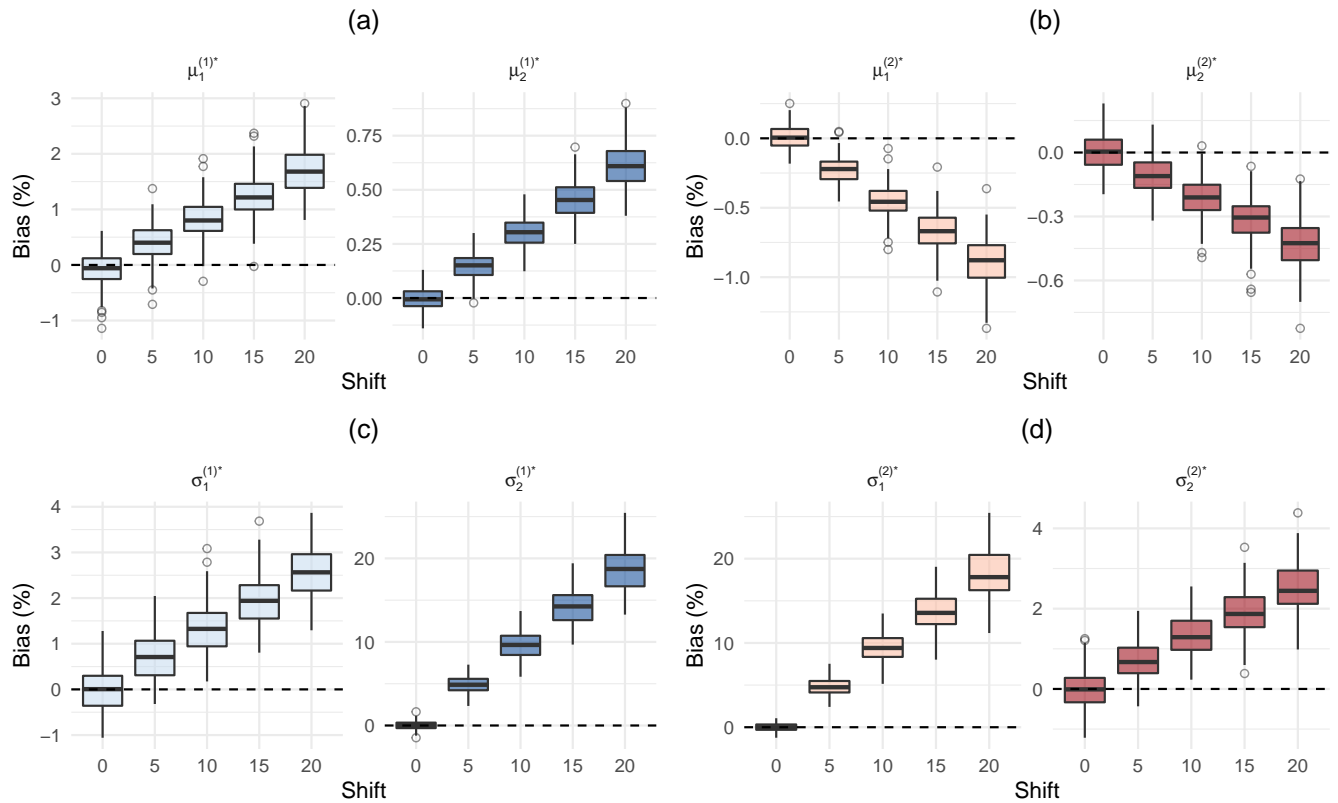


Fig. 4. Sample of results from the simulation experiment. Displayed is the percentage bias obtained across all 200 replications. The means of the accelerations under state i associated with fine-scale HMM k are denoted by $\mu_i^{(k)*}$ (panels (a) and (b)); the corresponding variances are denoted by $\sigma_i^{(k)*}$ (panels (c) and (d)). Full results are displayed in Section S2.2 of the Supplementary Material.

237 problematic even in basic HMMs (Pohle et al., 2017), can be exacerbated in HHMMs. In this example, it is
 238 likely a 4-state fine-scale process will be preferred: two to represent the true process and two to represent
 239 the unintended overlap of the two processes. This can cause misleading ecological conclusions about
 240 patterns in animal behaviour.

241 3.3 Opportunities and future prospects

242 Recent advances in bio-logging technology have led to the ability to track animals for increasingly long
 243 time periods at increasingly fine temporal resolutions (Kays et al., 2015; Lennox et al., 2017). As these
 244 data are often collected by multiple sensors with different sampling frequencies (e.g., GPS tags, dive
 245 loggers, or accelerometers), HHMMs provide a natural framework to jointly model these data and make

246 inference on the multi-scale nature of animal behaviour depicted by these new types of data. However, as
247 demonstrated in Section 3.2, the true data-generating processes underlying these multi-scale data do not
248 necessarily follow the assumed dependence structure and thus the model can misconstrue inference on
249 the underlying behavioural processes.

250 To best exploit the opportunities offered by multi-scale data using HHMMs, more flexible dependence
251 structures are needed. There are three possible approaches that future research could pursue: (1) estimate
252 the optimal resolution of the coarse-scale state process and compute coarse-scale observations on this
253 time interval, rather than have this be determined by the sampling protocol; (2) develop models where
254 switches between coarse-scale states and fine-scale HMMs is smoother and not a binary change; or (3)
255 allow the time spent in each coarse-scale state to depend on the fine-scale state-switching dynamics
256 and the fine-scale observations. If such extensions can overcome the problems outlined in this section,
257 HHMMs have the opportunity to become a promising tool to draw a complete picture of animal behaviour,
258 where an animal's movement decisions made at various time scales, ranging from seasonal migration
259 over diurnal activity to movements of individual body parts, can all be modelled in a joint modelling
260 framework.

261 **4 Random Effects**

262 **4.1 Introduction**

263 HMMs are often used to make inference on multiple time series of data where each arises from a different
264 sampling unit (e.g., movement data from different animals or abundance indices from different populations).
265 One aim in such studies is to elicit common patterns that are exhibited across sampling units, for example,
266 the relationship between an animal's movement and an environmental covariate.

267 It is common in this case to assume *complete pooling* of the parameters of the HMM for ease of
268 interpretability, that is, one assumes the state-dependent distributions, the state-switching dynamics, and
269 all parameters are the *same* across all sampling units. Nonetheless, ecological sampling units are often
270 heterogeneous and so models that do not account for this variation can be too inflexible and lead to invalid
271 inference. There are two alternatives to complete pooling: *no pooling*, where essentially one fits a separate
272 HMM to each sampling unit and forgoes the ability to infer common patterns statistically, or *partial pooling*.
273 Partial pooling refers to including random effects, either continuous-valued or discrete-valued (Zucchini
274 et al., 2017; McClintock, 2021), in the observation or state processes to account for the heterogeneity among
275 sampling units. McClintock (2021) conducted an in-depth simulation to demonstrate when inclusion of
276 random effects on the state-switching process would affect inference and state prediction. For this reason,
277 we focus on when random effects are included on the state-dependent distributions instead.

278 As an example, consider the garter snake movement data set in Leos-Barajas et al. (2017) where a
279 time series of observed distances moved was recorded for multiple snakes. Suppose *a priori* we expect
280 3 underlying states, but know the movements under these states may vary by snake. We can model
281 these data as a 3-state HMM where the state-dependent distributions are gamma distributions with mean
282 $\mu_{i,n}$ and standard deviation $\sigma_{i,n}$ for states $n = 1, 2, 3$ and snakes $i = 1, \dots, K$. For complete pooling,
283 one would assume $\mu_{1,n} = \mu_{2,n} = \dots = \mu_{K,n}$ for all n (and similarly for σ). For no pooling, one would
284 estimate each mean and standard deviation separately for each snake and state. For partial pooling, one
285 could assume a hierarchical model where there is a population-level parameter, e.g. μ_n for state n , and
286 each snake's individual-level parameter varies around this population-level mean with some standard
287 deviation τ_n for state n , i.e. $\mu_{k,n} \sim N(\mu_n, \tau_n)$. This is the simplest way to partially pool across individuals,
288 but it can lead to an unexpected pitfall: the added flexibility can overfit to certain individuals and make
289 interpretation of the underlying states at the population-level difficult.

290 4.2 Pitfall

291 At present, when including random effects in HMMs, two key assumptions are implicitly made (i) the K
292 individual time series exhibit the same number of distinct states N and (ii) the ecological interpretation of
293 the states is consistent across sampling units. That is, state n across all individuals should be comparable
294 and serve as a proxy for the same latent ecological behavior. However, in practice, there is seldom a way to
295 know *a priori* if all individuals exhibited the same number of states during the period of their observation
296 or if the states are estimable given the data collected. Random effect modelling, as currently developed for
297 HMMs, can account for individual heterogeneity, but does not have any theoretical grounding to prevent
298 their accommodation of this heterogeneity from deforming interpretation of the population-level states.
299 Even when the model assumed is correct, one may be unable to recover the true parameter values (see
300 Section S3 of the Supplementary Material).

301 This is a key pitfall in HMMs. The difficulties that arise when accounting for individual heterogeneity
302 in the state-dependent distributions stem from a lack of estimability and interpretability of the states
303 across individuals. Thus, as it stands, partial pooling may not allow researchers to understand individual
304 heterogeneity exhibited in the population of interest by simply incorporating random effects as this can
305 lead to inaccurate inference due to a lack of estimability of the state-dependent parameters.

306 For the garter snake movement example, Fig. 5 displays the estimated state-dependent distributions of
307 four snakes, along with 95% pointwise credible intervals, for the model with complete pooling and partial
308 pooling of the means. Estimated state-dependent distributions for all snakes are provided in Section S3 of
309 the Supplementary Material. Across both models, the population level estimates of the state-dependent
310 distributions demonstrate three distinct states (more details in Leos-Barajas et al. (2017)), yet there are
311 clear differences (especially in state 2) indicating the possible importance of accounting for individual

312 variation. However, the individual-specific state-dependent distributions in the partial pooling model
313 demonstrate crucial differences which are challenging to address. For example, for many individuals
314 (e.g. snakes 1 and 18) there is a large overlap between state 1 and state 2 distributions, e.g. $\mu_{k,2}$, for
315 $k \in \{1, \dots, K\}$, have lower bounds as small as 0.04, compared to a 95% credible interval for μ_2 of (0.11,
316 0.20). This raises a question: are we capturing true differences across states in individuals, or are we
317 over-fitting, or are the individual-specific parameters not estimable? At present, there is no mechanism in
318 random effect modelling for HMMs to address this question without simply making more assumptions.

319 **4.3 Opportunities**

320 Inclusion of random effects in a HMM provides opportunities to learn about individual heterogeneity,
321 personality, and preference within a given population of interest. However, we demonstrate that even
322 under correct model specification, a HMM with random effects in the observation process may not be able
323 to recover the true individual-specific state-dependent distributions and can lead to biased inferences for
324 the state-switching dynamics (see Section S3 of the Supplementary Material). As such, one opportunity is
325 to understand under which experimental designs a HMM with random effects is estimable.

326 In addition to tackling the issue of estimability, another opportunity is to develop methodology that
327 better respects the usual aim of using a HMM: to discern individual-specific state-dependent distributions
328 that are *different enough* from one another across states and are coherent with the population-level
329 inference. In Fig. 5, the densities for state 1 and 2 have more overlap than is present in the population level
330 estimates, even when the aim is to capture three *distinct* states. Enforcing this criteria may be possible
331 via selection of informative prior distributions, if inference is conducted in a Bayesian framework, or
332 appropriate regularization, constructed through the elicitation of domain expertise. For instance, one

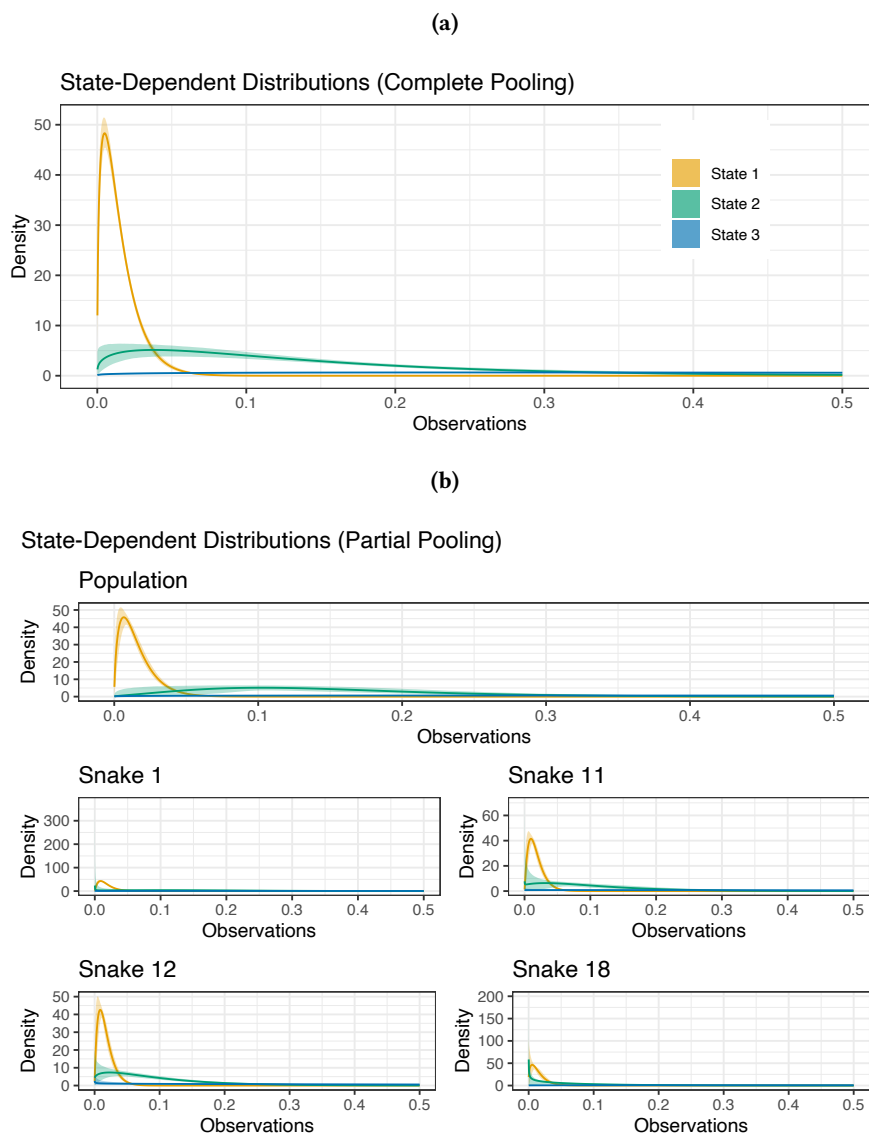


Fig. 5. Estimated state-dependent distributions, unweighted, for models with complete pooling (a) and partial pooling (b) of the state-dependent means, along with 95% pointwise credible intervals.

333 possibility is to assign an informative prior on the difference between $\mu_{n+1} - \mu_n$, for $n \in \{1, \dots, N - 1\}$,
 334 in order to enforce separation between densities.

335 If further research into the methodological development of HMMs with random effects proves to
 336 provide robust results for a variety of real-world data collection scenarios, it opens up the opportunity for

337 HMMs to provide insights into individual-specific movement dynamics and how distinct animal behaviors
338 manifest across a population.

339 **5 Continuous State Spaces**

340 In a HMM, the underlying state S_t takes one of a finite number of values; however, HMMs can be used
341 for *approximate* inference when S_t varies over an infinite number of states (Zucchini et al., 2017). Models
342 where S_t varies in a continuous state space are called “state space models” (SSMs) (Auger-Méthé et al.,
343 2021). The connection between state space models, hidden Markov models, and discretisation is well
344 known (Kitagawa, 1987; Anderson-Sprecher and Ledolter, 1991). In ecology, the approximate HMMs are
345 used to model animal movement (Pedersen et al., 2011), population dynamics (Besbeas and Morgan, 2019),
346 distance sampling sightings (Glennie et al., 2021), missing continuous covariates in capture-recapture
347 (Langrock and King, 2013), and for moving activity centres in spatial capture-recapture (Glennie et al.,
348 2019). Yet, the method remains an obscurity for many ecological statisticians, e.g., Patterson et al. (2017),
349 in a review of animal movement modelling, described it as “underutilized”.

350 In this section, we describe the ideas involved when approximating a SSM with a HMM, the current
351 pitfalls of this method, and the opportunities improving this method can bring to ecological applications.

352 **5.1 Example**

353 In this section, we provide an example of constructing an approximate HMM for a continuous-time,
354 state-switching animal movement SSM. This is to introduce the reader to the important ideas behind this
355 method. The idea is to break the continuous hidden state in the SSM into a discrete, binned state for a
356 HMM. We use continuous-time, state-switching animal movement with Brownian motion as an example
357 (Pedersen et al., 2011), where the data consist of observations of the animal’s location over (possibly
358 irregular) time. We assume an animal has two latent behavioural states and the diffusion rate depends

359 on its behaviour. The hidden variable in this SSM is the animal's location in 2D space coupled with its
360 behavioural state. Fig. 6 (Panel b) shows an example where 2D space has been split into a 5×2 grid.
361 The approximate HMM will have 20 hidden states in this case as each hidden state represents where the
362 animal is on the 2D grid and what behaviour the animal is in.

363 Once the states are defined for the HMM, the transition probability matrix or, if working in continuous
364 time, transition rate matrix is defined. The transition probabilities must be derived from the parameters
365 and state process in the SSM. If one can compute from the SSM the transition probabilities in continuous
366 space (as in Mews et al., 2020b) then one can compute the transition probabilities directly in the discrete
367 space. However, this is not always possible, e.g., with spatially-varying SSMs or state-switching SSMs.
368 Following Pedersen et al. (2011), we can derive a transition rate matrix for the HMM that mimics the
369 state process of the SSM using a partial differential equation (PDE; e.g. see Okubo and Levin, 2001). There
370 are several methods (e.g., finite differencing, finite volume, or finite element) to convert this PDE into a
371 transition matrix (Quarteroni and Valli, 2008). The parameters of the SSM determine the switching rates
372 for the approximating HMM. Fig. 6 (Panel a) shows an example of a derived transition rate matrix for a
373 behaviour-switching Brownian motion model on the 5×2 grid. Note that for most PDE methods the
374 derived transition matrix is *sparse* (i.e. most of its entries are zero).

375 Once the transition matrix over the grid is specified, one can fit the approximate HMM in the usual
376 way and estimate the parameters of the SSM. Fig. 6 (Panel c) shows two examples of using a derived
377 transition rate matrix to update the probability distribution over the 2D-behaviour space. Section S4
378 of the Supplementary Material provides a full example of building HMM approximations from PDEs,
379 including all code for constructing the necessary matrices and computing the likelihood.

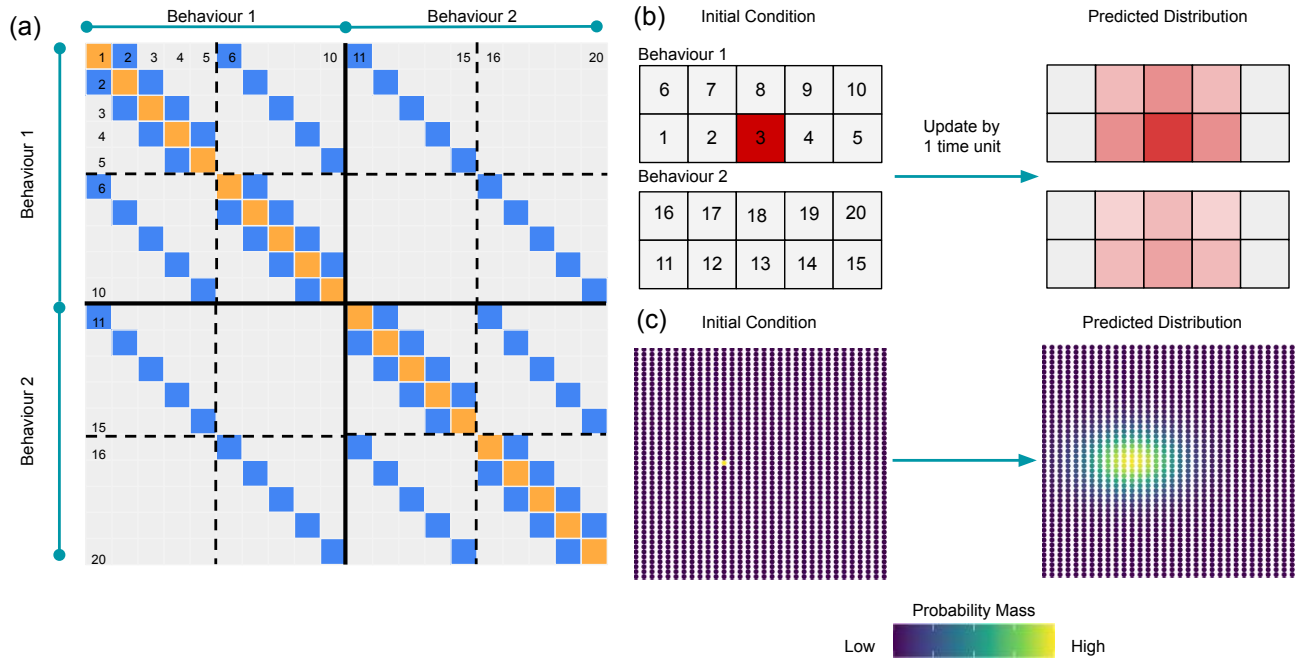


Fig. 6. Example of hidden Markov model for state-switching animal movement over 2D (x, y) space: (a) the transition rate matrix derived from the continuous-space, continuous-time partial differential equation model (Pedersen et al., 2011) where non-zero entries are coloured and some rows and columns are numbered to indicate what grid cell in 2D-behaviour space (Panel b) they refer to, solid lines demarcate blocks of the matrix corresponding to different behavioural states and dotted lines demarcate blocks corresponding to grid cells with different y values; (b) the 2D-behaviour space the transition rate matrix corresponds to with 2 behavioural states, each with 5 grid cells in the x direction and 2 in the y direction, each grid cell is numbered and has colouring to depict the initial distribution of an animal over this space (here the animal is in grid cell 3 in behaviour 1 with probability 1) and the predicted distribution of the animal’s location after 1 time unit (greater transparency indicates less probability mass); (c) a numerical example with 1 behavioural state using a higher grid resolution with an initial distribution of the animal in a single grid cell with probability 1 and the predicted distribution one time unit later under 2D Brownian motion, see Section S4 of the Supplementary Material for the code to compute this approximation.

380 5.2 Pitfalls

381 The biggest pitfall to using approximate HMMs to fit SSMs is the curse of dimensionality. For higher
382 dimensional hidden processes (e.g. where an animal's location and velocity are both hidden, a four-
383 dimensional space), the number of states in the HMM quickly becomes computationally infeasible. For
384 animal movement models, this inhibits the maximum number of behavioural states one can consider and
385 precludes the possibility of incorporating directional persistence. For population dynamics, this limits
386 the number of groups in age-structured models. Though this curse is unavoidable, future research could
387 alleviate it: irregular gridding of the latent space can focus inference on important parts (Pedersen and
388 Weng, 2013), only parts of the space can be updated when necessary, parallelised sparse matrix-vector
389 products and sparse matrix exponential methods can be more fully exploited (Sidje, 1998; Sherlock, 2021),
390 or sparse grids can reduce the number of states without compromising accuracy (Garcke, 2012). Many
391 of these proposed ideas for future research have direct analogy with extensions of simulation-based
392 methods.

393 The second pitfall is encountered when deciding whether to compute the transition probabilities directly
394 or by PDE. Both have limitations. The direct approach requires a known solution to the SSM and, for
395 large state-spaces, a threshold be set such that transition probabilities below that threshold are taken to be
396 zero as computing all pairwise transition probabilities is infeasible. The PDE approach has the advantage
397 that the transition rate matrix's sparsity is fixed no matter the time interval between observations and no
398 thresholding is necessary; however, the PDE approach is, at present, limited to SSM with Gaussian state
399 processes and requires one to compute a sparse matrix *exponential*. Future research is needed to expand
400 one or both of these approaches to a wider range of SSMs.

401 The final pitfall concerns the PDE approach in particular. The transition rate matrix is derived using PDE
402 techniques from applied mathematics. These techniques are designed to compute an accurate solution
403 to the PDE given the parameters are known. In ecological applications, however, the aim is different:
404 to compute the solution many times and to determine the optimal parameters for the PDE given the
405 data. Some techniques suitable for the former purpose may be subtly unsuitable for the latter, e.g., when
406 dealing with artificial diffusion or cross-diffusion (Quarteroni and Valli, 2008). In Section S4.4 of the
407 Supplementary Material, we show an example of this problem for directed animal movement. Future
408 research should consider the effect this may have on inference: most applications of the PDE method in
409 ecology have investigated the efficacy of state prediction and not parameter recovery (Gatti et al., 2021).

410 **5.3 Opportunities**

411 This method could have wide application in ecology if future research into its pitfalls led to improved
412 computational and statistical implementation. To show this, we consider three areas where this method
413 has shown some promise, but could be further exploited: animal movement, population dynamics, and
414 encounter modelling.

415 For animal movement, our example described one possible approach (Pedersen et al., 2008; Thygesen
416 et al., 2009; Pedersen et al., 2011). It continues to be used (Jonsen et al., 2013; Braun et al., 2018; Haase
417 et al., 2021), but only in limited contexts (focusing on light-based geolocation). Despite this, the flexibility
418 of the HMM approach offers solutions to many common modelling needs that arise with animal telemetry:
419 non-Gaussian, multivariate observations are easily accommodated, unlike common alternatives based
420 on Kalman filtering (Johnson et al., 2008); continuous-time behaviour-switching can be accounted for
421 and inference on where, not just when, behaviours are exhibited is possible (Pedersen et al., 2011), while
422 alternatives make it more difficult to make this connection (see Section 6); barriers and obstacles to

423 movement (e.g., land for marine animals) are trivially accounted for, while this remains difficult for
424 simulation-based techniques; and environmental covariates can drive step-selection in continuous time,
425 similar to continuous-time discrete-space models (Hanks et al., 2015), but with flexibility to include
426 behaviour-switching and link movement to a continuous-space model.

427 For population dynamics, De Valpine and Hastings (2002) introduced a discrete-time HMM approxima-
428 tion to general state-space population dynamics models. Besbeas and Morgan (2019) and Besbeas and
429 Morgan (2020) generalise the approach and make the connection to HMMs explicit. These approaches
430 differ from the example in Section 5.1: they do not consider continuous-time and compute transition
431 probabilities directly rather than via a PDE. The advantage of the HMM approach is the ability to model
432 nonlinear dynamics that the standard Kalman filter cannot accommodate. It also allows for multivariate
433 observations on population dynamics to be incorporated into a single model. Note that the state-switching
434 models, continuous-time modelling, or sparse matrix algorithms, all present in the animal movement
435 applications of this method, are yet to be fully explored for population dynamics models.

436 For encounter models, we refer to applications where animal movement is inferred from encounters
437 animals have with detectors, e.g., human observers, cameras, or acoustic devices. The approximate HMM
438 is easily extended to allow for custom detection models. As with telemetry data, these detections can
439 be used to infer individual animal movement (Pedersen and Weng, 2013; Dorazio and Price, 2019). Yet,
440 more powerfully, these individual animal movement models can be incorporated into population-level
441 models such as distance sampling or spatial capture-recapture (e.g. McClintock et al., 2021). In each of
442 these methods, a critical quantity is the probability *any* given animal in the population could be detected.
443 When animals can move, this detection probability is calculated by averaging over all the possible (yet
444 unobserved) paths an individual animal *could* have travelled. This is precisely the quantity the forward
445 algorithm for HMMs can be used to compute. This approach has been taken to incorporate animal

446 movement into both distance sampling (Glennie et al., 2021) and spatial capture-recapture (Glennie et al.,
447 2019). Furthermore, this methods provides the opportunity to build a general statistical approach to
448 encounter modelling (Gurarie and Ovaskainen, 2013).

449 **6 Spatial analysis and spatial interpretation of decoded states**

450 Inferring the behavioural state of an animal based on telemetry data is an increasingly popular application
451 of HMMs in ecology. Telemetry data are primarily made up of geographical locations, which are inherently
452 spatial, and HMMs for such data often characterise animal movement as a time series of steps and turns
453 (e.g. Langrock et al., 2012), either on the plane (2D) or in a volume (3D). Inferences from movement HMMs
454 can therefore assign behavioural states to spatial locations, which can be visualised in space by plotting
455 the predicted state assignments on a map. Under these basic HMMs, the model is blind to the spatial
456 mechanisms that give rise to certain behaviours. The spatial locations of the decoded states are therefore
457 irrelevant to the model, but highly relevant to the ecological application, and the ecological interpretation
458 of decoded states. A common extension to bring spatial mechanism into the model is to incorporate
459 spatial covariate effects on the transition probabilities or parameters of the state-dependent observation
460 distributions (e.g. Morales et al., 2004; Langrock et al., 2012; McClintock et al., 2012; Rivest et al., 2016; Mul
461 et al., 2020). This creates an explicit link between space and movement behaviour. The pitfall, however,
462 with this approach is that the appropriate structure for a given research question can quickly lead to a
463 model that is very complex, has a large number of parameters, and whose results are difficult to interpret.
464 For these reasons, an attractive alternative is a two-stage approach, where a relatively simple HMM is
465 first used to get state assignments for spatial locations, and post-hoc analyses are conducted to infer the
466 relationship between the states and the spatial context in which they occur (e.g. Breed et al., 2009; Nickel

467 et al., 2021). Though this is a simpler approach, there has been little research into how best to build a
468 statistically robust and rigorous two-stage model.

469 To illustrate the pitfall in using HMMs to make spatial inferences and spatially interpreting the decodes
470 states, we present a case study where HMMs were used to make inferences about animal movement from
471 tracking data, and secondary spatial analyses were carried out to address specific spatial, ecological and
472 behavioural questions.

473 **6.1 Pitfall example: Do hummingbirds use landmarks to remember spatial locations?**

474 The movement trajectories of fourteen rufous hummingbirds (*Selasphorus rufus*) were recorded in three-
475 dimensions during a field experiment (Westcastle Valley, Canadian Rockies, Alberta, Canada, May-July
476 2014) to investigate spatial memory and learning. The data (step length, pitch angle, yaw angle) were
477 analysed in (Pritchard et al., 2021) using a HMM, with distance to the location where a flower-shaped
478 feeder was previously situated, and the presence or absence of landmarks, as covariates on the transition
479 probability matrix.

480 The objective of the study was to gain a better understanding of how hummingbirds learn rewarding
481 spatial locations, and as such there was an interest in *where*, with respect to the feeder and landmarks,
482 certain movement behaviours occurred, as illustrated in Fig. 7. Knowing if birds were more likely to
483 display targeted searching behaviour in particular parts of space would allow inferences about what
484 specific features were learned and remembered. Hypothesis tests (t-tests) were carried out to determine
485 whether inferred searching state locations and stops made by experienced birds (with several prior visits
486 to the feeder in the presence of landmarks) were closer to the feeder location than would have been
487 expected by chance, compared to naive birds with a single prior visit. While this is a reasonable approach

488 and provides some insight into where behaviours occur, it does not fully utilise the information contained
 489 in the spatial distribution of states, or account for uncertainty in the state assignments.

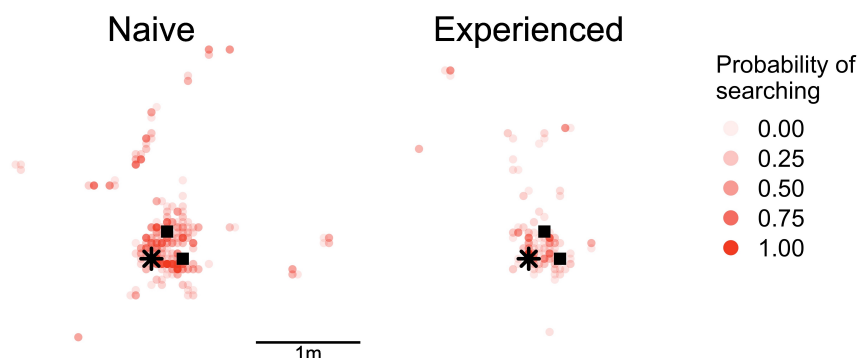


Fig. 7. The spatial distribution of locations assigned the searching state for naive birds, who only had one prior visit to the feeder (asterisk) before it was removed, and experienced birds who had several prior visits. The black squares mark the location of two artificial landmarks that were present during the experiment, while the feeder had been removed. The colour brightness gives the state probability, to convey uncertainty in state assignments.

490 6.2 Opportunities

491 This pitfall is at odds with what many, if not most, ecological and conservation science studies want to
 492 achieve when analysing telemetry data: an understanding of why behaviours occur where they do. In
 493 addition, the most easily implemented conservation and management interventions are often place-based,
 494 for example Marine Protected Areas. To inform place-based management decisions with inferences
 495 from movement HMMs, we must be able to identify both important habitat (e.g. stop-over sites) and
 496 the behaviour(s) associated with it (e.g. migration). This is key for identifying what parts of space have
 497 disproportionate functional value to an animal population and deserve further protection or management
 498 (Lennox et al., 2019). Extending the use of HMMs to better answer these questions or promoting alternative
 499 methods will have widespread impact on conservation practice, feeding into policy and management

500 decisions about place- and time-based conservation of wildlife, as well as their spatial and temporal
501 exposure to risk.

502 There are two questions any such proposed extension must address: (1) How can spatial inference
503 on behavioural state be made rigorously and robustly? (2) How can uncertainty in state be accounted
504 for? For HMMs, future research could consider a more refined two-stage approach (e.g. taking predicted
505 states, or, repeatedly drawing from the predicted state probabilities from a HMM and then performing
506 spatial analyses), implementing carefully thought-through spatio-temporal transition probability matrices
507 and/or observation distributions where interpretability permits, or exploiting the existing ability to
508 make spatial inference with continuous-space HMMs (as described in Section 5). In the hummingbird
509 example, the spatial research question could have instead been addressed within the HMM by including
510 an interaction term between level of experience and distance to the flower on the state transitions. It is
511 possible, however, that alternative, explicitly spatial methods, such as continuous-time discrete-space
512 models (Hanks et al., 2015) or Langevin diffusion models (Michelot et al., 2019), could be extended to
513 multiple behavioural states, thereby yielding state-specific utilization distributions that could inform
514 behaviour- and place-based management decisions. Such extensions could be embedded within the
515 continuous-time HMM framework by formulating the observation distribution accordingly, possibly
516 using a discrete-time approximation (e.g. Equation (S1) in Section S1.1 of the Supplementary Material), as
517 has been recently done (e.g. `momentuHMM` version 2.0.0; McClintock and Michelot, 2018).

518 **7 Conclusion**

519 Hidden Markov models are a versatile class of models that will continue to be developed for and applied
520 to ecological problems. In this review, we have highlighted capabilities of HMMs that could be fruitful for
521 ecological applications, but where one must be aware of current statistical and computational pitfalls.

522 HMMs are becoming more complex in structure to better realise the potential of the more heterogeneous,
523 multi-scale, multi-dimensional data being collected and the more detailed research questions being asked.
524 This can make HMMs a valuable tool to capture a wide variety of observations, to model multiple processes
525 acting at difference scales, to describe individual and population-level effects, and to link state-switching
526 to space as well as time. With these developments, however, come greater difficulties in interpreting these
527 models, justifying their assumptions, and fitting these models with current computational capabilities. We
528 hope this paper heightens awareness of the trouble more complex models can bring and what opportunities
529 for future solutions these problems can inspire.

530 **Acknowledgements**

531 The authors have no conflicts of interest. The findings and conclusions in the manuscript are those of the
532 author(s) and do not necessarily represent the views of the National Marine Fisheries Service, NOAA.
533 Any use of trade, product, or firm names does not imply an endorsement by the US Government. RG's
534 contribution was funded by the Biometrika Trust.

535 **Conflict of Interest**

536 None of the authors have a conflict of interest.

537 **Data Availability**

538 Garter snake data (used in Section 4) is available from this link: [https://link.springer.com/
539 article/10.1007/s13253-017-0282-9#Sec19](https://link.springer.com/article/10.1007/s13253-017-0282-9#Sec19).

540 **Author Contributions**

541 All authors contributed equally to the conception and implementation of the review. RG coordinated the
542 writing of the manuscript and all authors contributed substantially and critically to the manuscript. All
543 authors gave final approval for submission.

544 **Supplementary Material**

545 There is supplementary material available with this manuscript.

546 **References**

- 547 Adam, T., Griffiths, C. A., Leos-Barajas, V., Meese, E. N., Lowe, C. G., Blackwell, P. G., Righton, D., and
548 Langrock, R. (2019). Joint modelling of multi-scale animal movement data using hierarchical hidden
549 Markov models. *Methods in Ecology and Evolution*, 10(9):1536–1550.
- 550 Altman, R. M. (2007). Mixed hidden markov models: an extension of the hidden markov model to the
551 longitudinal data setting. *Journal of the American Statistical Association*, 102(477):201–210.
- 552 Anderson-Sprecher, R. and Ledolter, J. (1991). State-space analysis of wildlife telemetry data. *Journal of*
553 *the American Statistical Association*, 86(415):596–602.
- 554 Aquino-Baleytó, M., Leos-Barajas, V., Adam, T., Hoyos-Padilla, M., Santana-Morales, O., Galván-Magaña,
555 F., González-Armas, R., Lowe, C., Ketchum, J., and Villalobos-Ortiz, H. (2021). Diving deeper into the
556 underlying white shark behaviours at Guadalupe Island, Mexico. *Ecology and Evolution*, pages 1–18.
- 557 Auger-Méthé, M., Newman, K., Cole, D., Empacher, F., Gryba, R., King, A. A., Leos-Barajas, V.,
558 Mills Flemming, J., Nielsen, A., Petris, G., et al. (2021). A guide to state–space modeling of ecological
559 time series. *Ecological Monographs*.
- 560 Beketov, M. A. (2014). *HMMCont: Hidden Markov Model for Continuous Observations Processes*. R package
561 version 1.0.
- 562 Besbeas, P. and Morgan, B. J. T. (2019). Exact inference for integrated population modelling. *Biometrics*,
563 75(2):475–484.

- 564 Besbeas, P. and Morgan, B. J. T. (2020). A general framework for modeling population abundance data.
565 *Biometrics*, 76(1):281–292.
- 566 Blackwell, P. G. (2018). Integrated continuous-time hidden Markov models. *arXiv preprint arXiv:1807.11907*.
- 567 Blackwell, P. G., Niu, M., Lambert, M. S., and LaPoint, S. D. (2016). Exact Bayesian inference for animal
568 movement in continuous time. *Methods in Ecology and Evolution*, 7(2):184–195.
- 569 Borchers, D., Distiller, G., Foster, R., Harmsen, B., and Milazzo, L. (2014). Continuous-time spatially
570 explicit capture–recapture models, with an application to a jaguar camera-trap survey. *Methods in*
571 *Ecology and Evolution*, 5(7):656–665.
- 572 Braun, C. D., Galuardi, B., and Thorrold, S. R. (2018). HMMoce: An R package for improved geolocation of
573 archival-tagged fishes using a hidden Markov method. *Methods in Ecology and Evolution*, 9(5):1212–1220.
- 574 Breed, G. A., Jonsen, I. D., Myers, R. A., Bowen, W. D., and Leonard, M. L. (2009). Sex-specific, seasonal
575 foraging tactics of adult grey seals (*halichoerus grypus*) revealed by state–space analysis. *Ecology*,
576 90(11):3209–3221.
- 577 Bureau, A., Shiboski, S., and Hughes, J. P. (2003). Applications of continuous time hidden Markov models
578 to the study of misclassified disease outcomes. *Statistics in Medicine*, 22(3):441–462.
- 579 Choquet, R., Garnier, A., Awuve, E., and Besnard, A. (2017). Transient state estimation using continuous-
580 time processes applied to opportunistic capture–recapture data. *Ecological Modelling*, 361:157–163.
- 581 De Valpine, P. and Hastings, A. (2002). Fitting population models incorporating process noise and
582 observation error. *Ecological Monographs*, 72(1):57–76.

- 583 Descamps, S., Tarroux, A., Cherel, Y., Delord, K., Godø, O., Kato, A., Krafft, B., Lorentsen, S., Ropert-
584 Coudert, Y., Skaret, G., and Varpe, Ø. (2016a). Data from: At-sea distribution and prey selection of
585 Antarctic petrels and commercial krill fisheries.
- 586 Descamps, S., Tarroux, A., Cherel, Y., Delord, K., Godø, O. R., Kato, A., Krafft, B. A., Lorentsen, S.-H.,
587 Ropert-Coudert, Y., Skaret, G., et al. (2016b). At-sea distribution and prey selection of Antarctic petrels
588 and commercial krill fisheries. *PLoS one*, 11(8):e0156968.
- 589 Dorazio, R. M. and Price, M. (2019). State-space models to infer movements and behavior of fish detected
590 in a spatial array of acoustic receivers. *Canadian Journal of Fisheries and Aquatic Sciences*, 76(4):543–550.
- 591 Fine, S., Singer, Y., and Tishby, N. (1998). The hierarchical hidden Markov model: analysis and applications.
592 *Machine Learning*, 32(1):41–62.
- 593 Garcke, J. (2012). Sparse grids in a nutshell. In *Sparse grids and Applications*, pages 57–80. Springer.
- 594 Gatti, P., Fisher, J. A., Cyr, F., Galbraith, P. S., Robert, D., and Le Bris, A. (2021). A review and tests of
595 validation and sensitivity of geolocation models for marine fish tracking. *Fish and Fisheries*, 22(5):1041–
596 1066.
- 597 Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A., and Rubin, D. B. (2013). *Bayesian data*
598 *analysis*. CRC press.
- 599 Glennie, R., Borchers, D. L., Murchie, M., Harmsen, B. J., and Foster, R. J. (2019). Open population maximum
600 likelihood spatial capture-recapture. *Biometrics*, 75(4):1345–1355.
- 601 Glennie, R., Buckland, S. T., Langrock, R., Gerrodette, T., Ballance, L., Chivers, S., and Scott, M. (2021).
602 Incorporating animal movement into distance sampling. *Journal of the American Statistical Association*,
603 116(533):107–115.

- 604 Gurarie, E. and Ovaskainen, O. (2013). Towards a general formalization of encounter rates in ecology.
605 *Theoretical Ecology*, 6(2):189–202.
- 606 Haase, S., Krumme, U., Gräwe, U., Braun, C. D., and Temming, A. (2021). Validation approaches of a
607 geolocation framework to reconstruct movements of demersal fish equipped with data storage tags in a
608 stratified environment. *Fisheries Research*, 237:105884.
- 609 Hanks, E. M., Hooten, M. B., and Alldredge, M. W. (2015). Continuous-time discrete-space models for
610 animal movement. *The Annals of Applied Statistics*, 9(1):145–165.
- 611 Jackson, C. H. (2011). Multi-state models for panel data: The msm package for R. *Journal of Statistical*
612 *Software*, 38(8):1–29.
- 613 Jackson, C. H., Sharples, L. D., Thompson, S. G., Duffy, S. W., and Couto, E. (2003). Multistate Markov
614 models for disease progression with classification error. *Journal of the Royal Statistical Society: Series D*
615 *(The Statistician)*, 52(2):193–209.
- 616 Johnson, D. S., London, J. M., Lea, M.-A., and Durban, J. W. (2008). Continuous-time correlated random
617 walk model for animal telemetry data. *Ecology*, 89(5):1208–1215.
- 618 Jonsen, I., Basson, M., Bestley, S., Bravington, M., Patterson, T., Pedersen, M. W., Thomson, R., Thygesen,
619 U. H., and Wotherspoon, S. (2013). State-space models for bio-loggers: A methodological road map.
620 *Deep Sea Research Part II: Topical Studies in Oceanography*, 88:34–46.
- 621 Kays, R., Crofoot, M. C., Jetz, W., and Wikelski, M. (2015). Terrestrial animal tracking as an eye on life
622 and planet. *Science*, 348(6240).
- 623 Kitagawa, G. (1987). Non-Gaussian state–space modeling of nonstationary time series. *Journal of the*
624 *American statistical association*, 82(400):1032–1041.

- 625 Langrock, R. and King, R. (2013). Maximum likelihood estimation of mark–recapture–recovery models in
626 the presence of continuous covariates. *The Annals of Applied Statistics*, 7(3):1709–1732.
- 627 Langrock, R., King, R., Matthiopoulos, J., Thomas, L., Fortin, D., and Morales, J. M. (2012). Flexible
628 and practical modeling of animal telemetry data: hidden Markov models and extensions. *Ecology*,
629 93(11):2336–2342.
- 630 Langrock, R. and Zucchini, W. (2011). Hidden markov models with arbitrary state dwell-time distributions.
631 *Computational Statistics & Data Analysis*, 55(1):715–724.
- 632 Lennox, R. J., Aarestrup, K., Cooke, S. J., Cowley, P. D., Deng, Z. D., Fisk, A. T., Harcourt, R. G., Heupel, M.,
633 Hinch, S. G., Holland, K. N., et al. (2017). Envisioning the future of aquatic animal tracking: technology,
634 science, and application. *BioScience*, 67(10):884–896.
- 635 Lennox, R. J., Engler-Palma, C., Kowarski, K., Filous, A., Whitlock, R., Cooke, S. J., and Auger-Méthé, M.
636 (2019). Optimizing marine spatial plans with animal tracking data. *Canadian Journal of Fisheries and*
637 *Aquatic Sciences*, 76(3):497–509.
- 638 Leos-Barajas, V., Gangloff, E. J., Adam, T., Langrock, R., Van Beest, F. M., Nabe-Nielsen, J., and Morales,
639 J. M. (2017). Multi-scale modeling of animal movement and general behavior data using hidden Markov
640 models with hierarchical structures. *Journal of Agricultural, Biological and Environmental Statistics*,
641 22(3):232–248.
- 642 Liu, Y.-Y., Li, S., Li, F., Song, L., and Rehg, J. M. (2015). Efficient learning of continuous-time hidden Markov
643 models for disease progression. *Advances in Neural Information Processing Systems*, 28:3500–3607.
- 644 MacKenzie, D. I., Nichols, J. D., Royle, J. A., Pollock, K. H., Bailey, L., and Hines, J. E. (2018). *Occupancy*
645 *Estimation and Modeling: Inferring Patterns and Dynamics of Species Occurrence*. Elsevier, 2nd edition.

- 646 McClintock, B. T. (2017). Incorporating telemetry error into hidden Markov models of animal movement
647 using multiple imputation. *Journal of Agricultural, Biological and Environmental Statistics*, 22(3):249–269.
- 648 McClintock, B. T. (2021). Worth the effort? A practical examination of random effects in hidden Markov
649 models for animal telemetry data. *Methods in Ecology and Evolution*, 12(8):1475–1497.
- 650 McClintock, B. T., Abrahms, B., Chandler, R. B., Conn, P. B., Converse, S. J., Emmet, R., Gardner, B.,
651 Hostetter, N. J., and Johnson, D. S. (2021). An integrated path for spatial capture-recapture and animal
652 movement modeling. *Ecology*, Accepted Author Manuscript e03473. <https://doi.org/10.1002/ecy.3473>.
- 653 McClintock, B. T., Johnson, D. S., Hooten, M. B., Ver Hoef, J. M., and Morales, J. M. (2014). When to be
654 discrete: the importance of time formulation in understanding animal movement. *Movement Ecology*,
655 2(1):1–14.
- 656 McClintock, B. T., King, R., Thomas, L., Matthiopoulos, J., McConnell, B. J., and Morales, J. M. (2012).
657 A general discrete-time modeling framework for animal movement using multistate random walks.
658 *Ecological Monographs*, 82(3):335–349.
- 659 McClintock, B. T., Langrock, R., Gimenez, O., Cam, E., Borchers, D. L., Glennie, R., and Patterson, T. A.
660 (2020). Uncovering ecological state dynamics with hidden Markov models. *Ecology Letters*, 23(12):1878–
661 1903.
- 662 McClintock, B. T. and Michelot, T. (2018). momentuHMM: R package for generalized hidden Markov
663 models of animal movement. *Methods in Ecology and Evolution*, 9(6):1518–1530.
- 664 McLachlan, G. J. and Krishnan, T. (2007). *The EM algorithm and extensions*. John Wiley & Sons.
- 665 Mews, S., Langrock, R., King, R., and Quick, N. (2020a). Continuous-time multi-state capture-recapture
666 models. *arXiv preprint arXiv:2002.10997*.

- 667 Mews, S., Langrock, R., Ötting, M., Yaqine, H., and Reinecke, J. (2020b). Maximum approximate likelihood
668 estimation of general continuous-time state-space models. *arXiv preprint arXiv:2010.14883*.
- 669 Michelot, T. and Blackwell, P. G. (2019). State-switching continuous-time correlated random walks.
670 *Methods in Ecology and Evolution*, 10(5):637–649.
- 671 Michelot, T., Gloaguen, P., Blackwell, P. G., and Étienne, M.-P. (2019). The Langevin diffusion as a
672 continuous-time model of animal movement and habitat selection. *Methods in Ecology and Evolution*,
673 10(11):1894–1907.
- 674 Morales, J. M., Haydon, D. T., Frair, J., Holsinger, K. E., and Fryxell, J. M. (2004). Extracting more out of
675 relocation data: Building movement models as mixtures of random walks. *Ecology*, 85(9):2436–2445.
- 676 Mul, E., Blanchet, M.-A., McClintock, B. T., Grecian, W. J., Biuw, M., and Rikardsen, A. (2020). Killer
677 whales are attracted to herring fishing vessels. *Marine Ecology Progress Series*, 652:1–13.
- 678 Nickel, B. A., Suraci, J. P., Nisi, A. C., and Wilmers, C. C. (2021). Energetics and fear of humans constrain
679 the spatial ecology of pumas. *Proceedings of the National Academy of Sciences*, 118(5):e2004592118.
- 680 Okubo, A. and Levin, S. A. (2001). *Diffusion and ecological problems: modern perspectives*. Springer.
- 681 Patterson, T. A., Parton, A., Langrock, R., Blackwell, P. G., Thomas, L., and King, R. (2017). Statistical
682 modelling of individual animal movement: an overview of key methods and a discussion of practical
683 challenges. *AStA Advances in Statistical Analysis*, 101(4):399–438.
- 684 Pedersen, M. W., Patterson, T. A., Thygesen, U. H., and Madsen, H. (2011). Estimating animal behavior
685 and residency from movement data. *Oikos*, 120(9):1281–1290.

- 686 Pedersen, M. W., Righton, D., Thygesen, U. H., Andersen, K. H., and Madsen, H. (2008). Geolocation of
687 North Sea cod (*Gadus morhua*) using hidden markov models and behavioural switching. *Canadian*
688 *Journal of Fisheries and Aquatic Sciences*, 65(11):2367–2377.
- 689 Pedersen, M. W. and Weng, K. C. (2013). Estimating individual animal movement from observation
690 networks. *Methods in Ecology and Evolution*, 4(10):920–929.
- 691 Plummer, M. (2017). *JAGS: Just Another Gibbs Sampler*. Version 4.0.3 User Manual.
- 692 Pohle, J., Langrock, R., van Beest, F. M., and Schmidt, N. M. (2017). Selecting the number of states in
693 hidden Markov models: pragmatic solutions illustrated using animal movement. *Journal of Agricultural,*
694 *Biological and Environmental Statistics*, 22(3):270–293.
- 695 Pritchard, D. J., Hurly, T. A., Photopoulou, T., and Healy, S. D. (2021). 3d flightpaths reveal the development
696 of spatial memory in wild hummingbirds. *In prep*.
- 697 Quarteroni, A. and Valli, A. (2008). *Numerical approximation of partial differential equations*. Springer
698 Science & Business Media.
- 699 R Core Team (2020). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical
700 Computing, Vienna, Austria.
- 701 Rivest, L.-P., Duchesne, T., Nicosia, A., and Fortin, D. (2016). A general angular regression model for
702 the analysis of data on animal movement in ecology. *Journal of the Royal Statistical Society: Series C*
703 *(Applied Statistics)*, 65(3):445–463.
- 704 Royle, J. A. and Kéry, M. (2007). A Bayesian state-space formulation of dynamic occupancy models.
705 *Ecology*, 88(7):1813–1823.
- 706 Rubin, D. B. (2004). *Multiple imputation for nonresponse in surveys*. John Wiley & Sons.

- 707 Sacchi, G. and Swallow, B. (2021). Toward efficient Bayesian approaches to inference in hierarchical
708 hidden markov models for inferring animal behavior. *Frontiers in Ecology and Evolution*, 9:249.
- 709 Sherlock, C. (2021). Direct statistical inference for finite markov jump processes via the matrix exponential.
710 *Computational Statistics*.
- 711 Sidje, R. B. (1998). Expokit: A software package for computing matrix exponentials. *ACM Transactions on*
712 *Mathematical Software (TOMS)*, 24(1):130–156.
- 713 Sidrow, E., Heckman, N., Fortune, S. M., Trites, A. W., Murphy, I., and Auger-Méthé, M. (2021). Modelling
714 multi-scale state-switching functional data with hidden markov models. *arXiv preprint arXiv:2101.03268*.
- 715 Thygesen, U. H., Pedersen, M. W., and Madsen, H. (2009). Geolocating fish using hidden Markov models
716 and data storage tags. In *Tagging and Tracking of Marine Animals with Electronic Devices*, pages 277–293.
717 Springer.
- 718 Zucchini, W., MacDonald, I. L., and Langrock, R. (2017). *Hidden Markov models for time series: an*
719 *introduction using R, Second Edition*. CRC press.