

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

14th June 2022

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

44 $S_t = i$. Key properties of HMMs are that they commonly deal with observations in regular, discrete
45 time; the state, S_t , switches between a pre-specified finite number of possible values; and there are
46 computationally efficient 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.

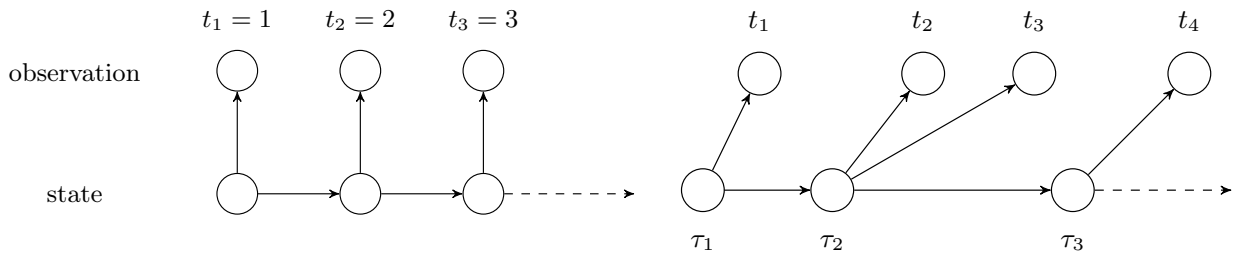


Fig. 1. Comparison of dependence graphs for discrete-time (left) and continuous-time (right) HMMs. In both cases, the distribution of an observation only depends on the current value of the state. The key difference is that, in a continuous-time HMM, the times of state transitions (τ_1, τ_2, \dots) and 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.,
74 2003; Liu et al., 2015). However, they remain rare in ecological studies, despite their broad applicability to
75 irregular data sets. We posit that they have been underutilised because the mathematical theory behind
76 continuous-time Markov processes is less intuitive than its counterpart in discrete time, and because of
77 the apparent lack of accessible software to readily apply this method. However, much of the inferential
78 framework developed for discrete-time HMMs can also be used in continuous time, based on the rela-
79 tionship between transition rates and transition probabilities in Equation (1). This includes the forward
80 algorithm and the Viterbi algorithm (Zucchini et al., 2017). In many situations, the implementation of
81 a continuous-time HMM is virtually identical to that of a discrete-time HMM, with the only difference
82 being that the likelihood function is parametrised in terms of transition rates. Yet, it is once you move
83 from discrete to continuous time that a central property of HMMs becomes less intuitive: the snapshot
84 property.

85 **2.2 Pitfall: snapshot property violation**

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

94 One context where the snapshot property is usually violated is state-switching models based on rel-
95 atively 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
110 then assessed the performance of the approximation based on bias between true and estimated para-
111 meters, and on the proportion of correctly-estimated states, for different time intervals of observation.
112 The simulation procedure is described in Section S1.2 of the Supplementary Material, and the results are
113 shown in Fig. 2. As expected, the approximation error increases with the mean time interval because
114 more switches can occur between observations (but each observation is assumed to only depend on the
115 state active at that time).

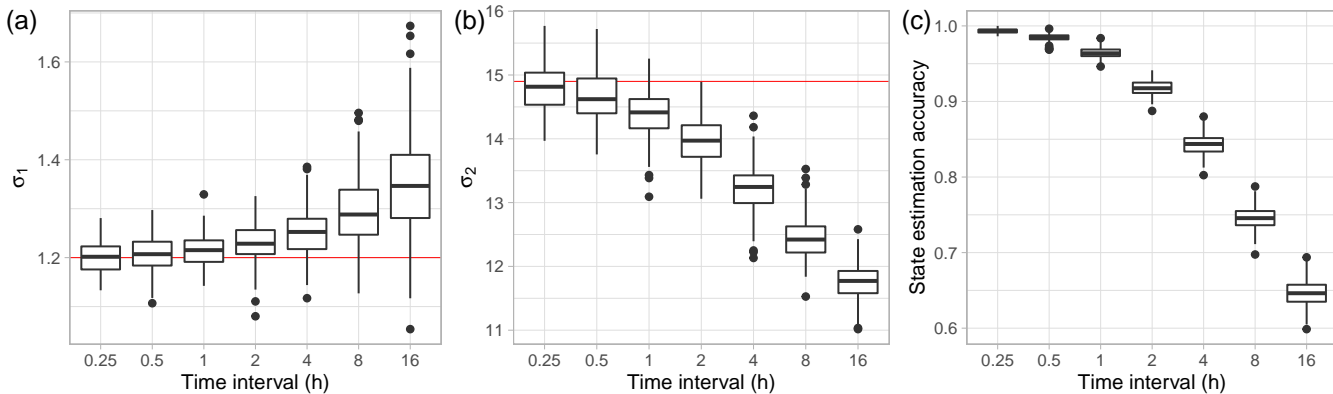


Fig. 2. Results of continuous-time HMM simulation study. Estimated parameters σ_1 (a) and σ_2 (b), and state estimate accuracy (c) for simulated data with different mean time intervals, from 0.25h to 16h. Each box contains 200 replications, for each simulation scenario.

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) rel-
 118 ative to the time scale of the observations. Our results support an intuitive rule of thumb for determining
 119 whether or not any given sampling interval may reasonably satisfy the snapshot property:

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

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 state-
 122 dependent observation distribution overlap, serial correlation in the hidden state sequence, and other
 123 properties of the data (e.g. Zucchini et al., 2017; McClintock, 2021). These results suggest that continuous-
 124 time HMMs should not be applied naively in ecological studies where the snapshot property is known
 125 to be violated, in particular when the time intervals of observations are long relative to the time scale of
 126 the state process.

127 **2.3 Opportunities**

128 Our simulation results indicated that it is important to consider whether or not the snapshot property
129 is 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
141 ensuring time intervals are sufficiently short relative to the transition rates), there are existing rem-
142 edies. One option is to augment the data with missing observation times (thereby yielding sufficiently
143 short intervals) and then integrate over the missing observations during model fitting. This integration
144 is relatively straightforward in Bayesian analysis (e.g., Gelman et al., 2013) and maximum likelihood
145 analyses using expectation-maximization algorithms (e.g., McLachlan and Krishnan, 2007). When the
146 forward algorithm is used to maximize the likelihood directly, multiple imputation methods can account
147 for the missing observations (e.g. Rubin, 2004; McClintock, 2017). These approaches are approximate
148 but relatively easy to implement. The challenge with exact methods is that the state transition times are

149 unknown. An exact but relatively difficult method to implement is the integrated continuous-time HMM
150 (Blackwell, 2018), which is a Bayesian approach that utilizes the efficient forward algorithm to estimate
151 the whole state process together with the model parameters. Further development of exact methods that
152 efficiently 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
156 a 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 a
183 different time scale: (1) a coarse-scale (e.g. hours) hidden process S_t for coarse-scale time t and (2) a fine-
184 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-
188 scale HMMs. The coarse-scale state that is active determines which fine-scale HMM models the fine-
189 scale observations. As the fine-scale observations do not only depend on the fine-scale state process but
190 also 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).

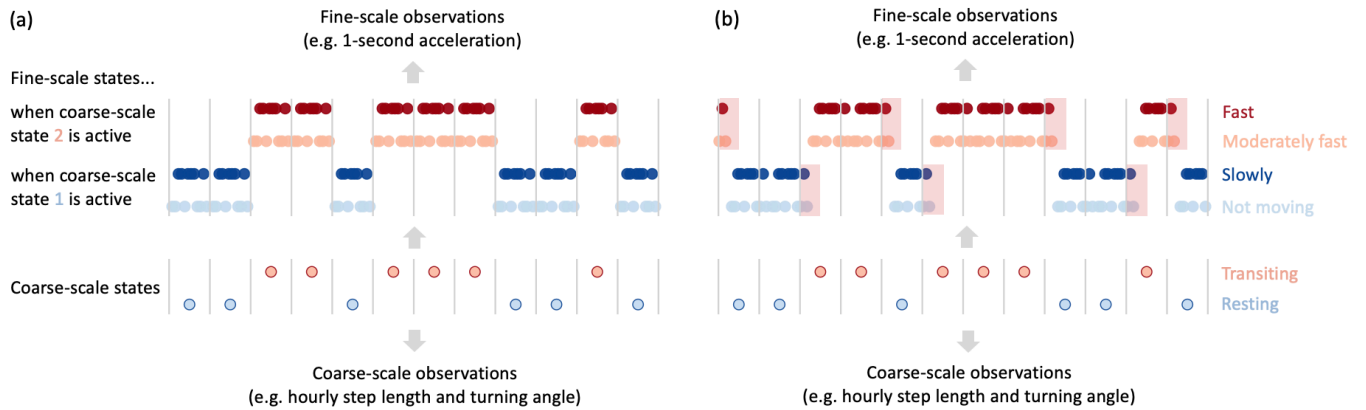


Fig. 3. Illustration of the state processes of an HHMM where the processes are aligned (a) and where this assumption is violated (b). A typical example for the latter scenario is an animal that performs (moderately) fast movements during the first few minutes of an hour that is, overall, characterised 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.

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,
 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

hour that is, overall, characterised by transiting behaviour (see the red-shaded areas in Fig. 3 (b) for an illustration of such a scenario). Similarly, the fine-scale observations are not necessarily driven by the *same* coarse-scale state process as the coarse-scale observations. Accelerations can be driven by other behaviours than resting and transiting, which drive step lengths and turning angles. In both examples, multi-scale data do not follow the dependence structure that is assumed by HHMMs.

To show the consequences of such a violation of the dependence structure, we conducted a simulation experiment (further details on the simulation procedure are provided in Section S2.1 of the Supplementary Material). Over 200 replications, we simulated a 2-state coarse-scale process on the hourly scale with 1000 observations of step length and turning angle, and a 2-state fine-scale process with 100 observations of acceleration. We then progressively shifted the fine-scale process by 0, 5, 10, 15, and 20 observations and computed the percentage bias in parameter estimates (full descriptions of the parameters are provided in Section S2.1 of the Supplementary Material). While we used deterministic shifts, in practice, they can also be probabilistically. In that regard, the deterministic shifts used for the simulation experiment can be thought of as being exemplary of a probabilistic shifting process with mean equal to the deterministic shifts and small variance. For probabilistic shifting processes with large variance, or scenarios where the fine-scale HMMs change probabilistically within a coarse-scale state, we expect that HHMMs often fail to infer distinct behavioural modes at the fine scale, as all fine-scale behaviours occur within all coarse-scale states, i.e., the two processes are less correlated. In such cases, separate HMMs for the two processes should be preferred over HHMMs. All models were fit using the R package `momentuHMM` (McClintock and Michelot, 2018).

Example results are displayed in Fig. 4 (full results are displayed in Section S2.2 of the Supplementary Material). Clearly, when the observations were simulated as assumed by the HHMM (i.e., without shifting the fine-scale process), the estimates are unbiased; however, the question is how severe the

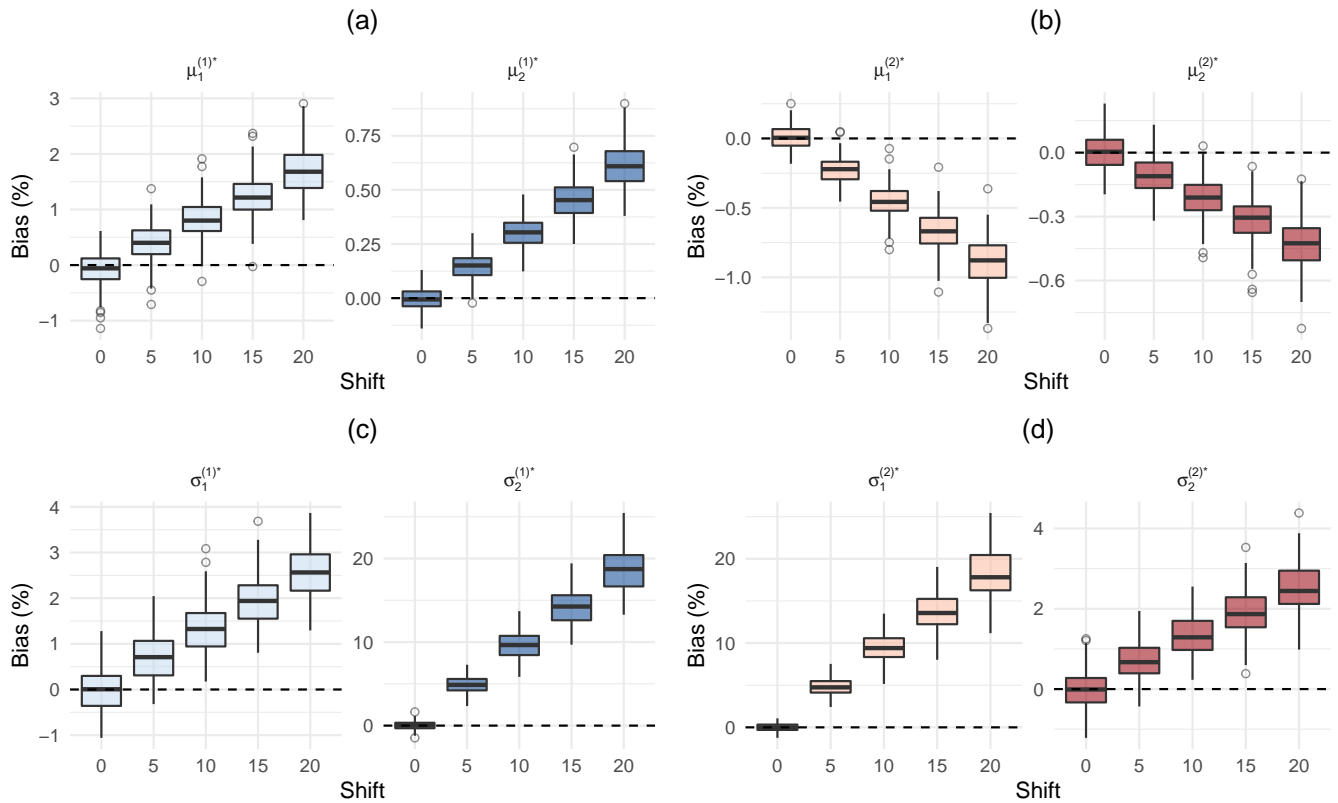


Fig. 4. Sample of results from the simulation experiment. Displayed is the percentage bias obtained across all 200 replications. The means of the accelerations under state i associated with fine-scale HMM k are denoted by $\mu_i^{(k)*}$ (panels (a) and (b)); the corresponding variances are denoted by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

229 bias increases as the fine-scale state process is shifted. While the bias remains relatively small for small
 230 shifts, it increases sharply with shifting of the fine-scale process. This severe bias is due to the fact that
 231 each of the two fine-scale HMMs must accommodate observations within each hour that truly belong to
 232 the alternate 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

237 problematic even in basic HMMs (Pohle et al., 2017), can be exacerbated in HHMMs. In this example,
238 it is likely a 4-state fine-scale process will be preferred: two to represent the true process and two to
239 represent the unintended overlap of the two processes. This can cause misleading ecological conclusions
240 about 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,
247 as demonstrated in Section 3.2, the true data-generating processes underlying these multi-scale data do
248 not necessarily follow the assumed dependence structure and thus the model can misconstrue inference
249 on 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 be-
258 haviour, where an animal's movement decisions made at various time scales, ranging from seasonal

259 migration over diurnal activity to movements of individual body parts, can all be modelled in a joint
260 modelling 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 dif-
264 ferent sampling unit (e.g., movement data from different animals or abundance indices from different
265 populations). One aim in such studies is to elicit common patterns that are exhibited across sampling
266 units, for example, 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,
269 and all parameters are the *same* across all sampling units. Nonetheless, ecological sampling units are
270 often heterogeneous and so models that do not account for this variation can be too inflexible and lead
271 to invalid inference. There are two alternatives to complete pooling: *no pooling*, where essentially one
272 fits a separate HMM to each sampling unit and forgoes the ability to infer common patterns statistic-
273 ally, or *partial pooling*. Partial pooling refers to including random effects, either continuous-valued or
274 discrete-valued (Zucchini et al., 2017; McClintock, 2021), in the observation or state processes to account
275 for the heterogeneity among sampling units. McClintock (2021) conducted an in-depth simulation to
276 demonstrate when inclusion of random effects on the state-switching process would affect inference and
277 state prediction. For this reason, we focus on when random effects are included on the state-dependent
278 distributions instead.

279 As an example, consider the garter snake movement data set in Leos-Barajas et al. (2017) where a
280 time series of observed distances moved was recorded for multiple snakes. Suppose *a priori* we expect 3

281 underlying states, but know the movements under these states may vary by snake. We can model these
282 data as a 3-state HMM where the state-dependent distributions are gamma distributions with mean $\mu_{i,n}$
283 and standard deviation $\sigma_{i,n}$ for states $n = 1, 2, 3$ and snakes $i = 1, \dots, K$. For complete pooling, one
284 would assume $\mu_{1,n} = \mu_{2,n} = \dots = \mu_{K,n}$ for all n (and similarly for σ). For no pooling, one would
285 estimate each mean and standard deviation separately for each snake and state. For partial pooling, one
286 could assume a hierarchical model where there is a population-level parameter, e.g. μ_n for state n , and
287 each snake's individual-level parameter varies around this population-level mean with some standard
288 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,
289 but it can lead to an unexpected pitfall: the added flexibility can overfit to certain individuals and make
290 interpretation of the underlying states at the population-level difficult.

291 4.2 Pitfall

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

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

307 For the garter snake movement example, Fig. 5 displays the estimated state-dependent distributions of
308 four snakes, along with 95% pointwise credible intervals, for the model with complete pooling and partial
309 pooling of the means. Estimated state-dependent distributions for all snakes are provided in Section S3 of
310 the Supplementary Material. Across both models, the population level estimates of the state-dependent
311 distributions demonstrate three distinct states (more details in Leos-Barajas et al. (2017)), yet there are
312 clear differences (especially in state 2) indicating the possible importance of accounting for individual
313 variation. However, the individual-specific state-dependent distributions in the partial pooling model
314 demonstrate crucial differences which are challenging to address. For example, for many individuals
315 (e.g. snakes 1 and 18) there is a large overlap between state 1 and state 2 distributions, e.g. $\mu_{k,2}$, for
316 $k \in \{1, \dots, K\}$, have lower bounds as small as 0.04, compared to a 95% credible interval for μ_2 of (0.11,
317 0.20). This raises a question: are we capturing true differences across states in individuals, or are we
318 over-fitting, or are the individual-specific parameters not estimable? At present, there is no mechanism
319 in random effect modelling for HMMs to address this question without simply making more assumptions.

320 **4.3 Opportunities**

321 Inclusion of random effects in a HMM provides opportunities to learn about individual heterogeneity,
322 personality, and preference within a given population of interest. However, we demonstrate that even
323 under correct model specification, a HMM with random effects in the observation process may not be able

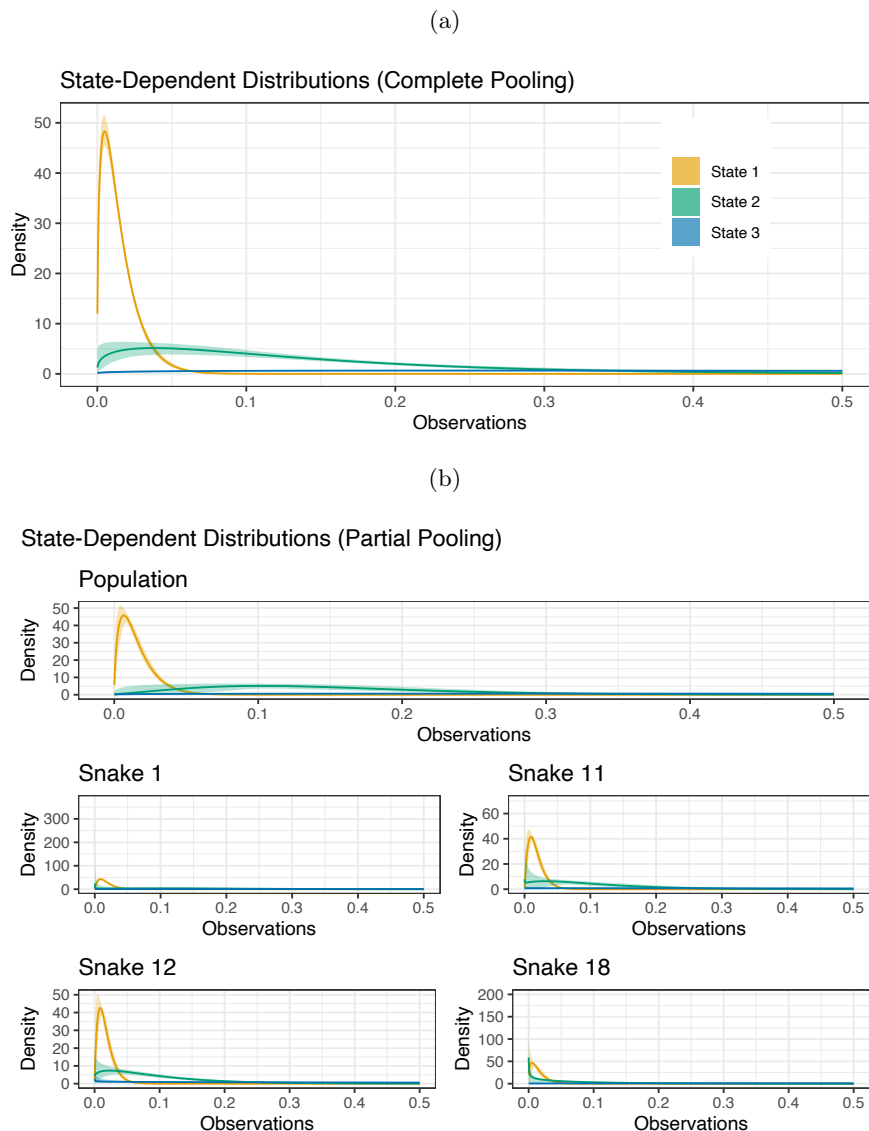


Fig. 5. Estimated state-dependent distributions, unweighted, for models with complete pooling (a) and partial pooling (b) of the state-dependent means, along with 95% pointwise credible intervals.

324 to recover the true individual-specific state-dependent distributions and can lead to biased inferences for
 325 the state-switching dynamics (see Section S3 of the Supplementary Material). As such, one opportunity
 326 is to understand under which experimental designs a HMM with random effects is estimable.

327 In addition to tackling the issue of estimability, another opportunity is to develop methodology that
 328 better respects the usual aim of using a HMM: to discern individual-specific state-dependent distributions

329 that are *different enough* from one another across states and are coherent with the population-level
330 inference. In Fig. 5, the densities for state 1 and 2 have more overlap than is present in the population level
331 estimates, even when the aim is to capture three *distinct* states. Enforcing this criteria may be possible
332 via selection of informative prior distributions, if inference is conducted in a Bayesian framework, or
333 appropriate regularization, constructed through the elicitation of domain expertise. For instance, one
334 possibility is to assign an informative prior on the difference between $\mu_{n+1} - \mu_n$, for $n \in \{1, \dots, N - 1\}$,
335 in order to enforce separation between densities.

336 If further research into the methodological development of HMMs with random effects proves to
337 provide robust results for a variety of real-world data collection scenarios, it opens up the opportunity
338 for HMMs to provide insights into individual-specific movement dynamics and how distinct animal
339 behaviors manifest across a population.

340 5 Continuous State Spaces

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

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

353 **5.1 Example**

354 In this section, we provide an example of constructing an approximate HMM for a continuous-time,
355 state-switching animal movement SSM. This is to introduce the reader to the important ideas behind
356 this method. The idea is to break the continuous hidden state in the SSM into a discrete, binned state
357 for a HMM. We use continuous-time, state-switching animal movement with Brownian motion as an
358 example (Pedersen et al., 2011), where the data consist of observations of the animal's location over
359 (possibly irregular) time. We assume an animal has two latent behavioural states and the diffusion rate
360 depends on its behaviour. The hidden variable in this SSM is the animal's location in 2D space coupled
361 with its behavioural state. Fig. 6 (Panel b) shows an example where 2D space has been split into a 5×2
362 grid. The approximate HMM will have 20 hidden states in this case as each hidden state represents where
363 the animal is on the 2D grid and what behaviour the animal is in.

364 Once the states are defined for the HMM, the transition probability matrix or, if working in continuous
365 time, transition rate matrix is defined. The transition probabilities must be derived from the parameters
366 and state process in the SSM. If one can compute from the SSM the transition probabilities in continuous
367 space (as in Mews et al., 2020b) then one can compute the transition probabilities directly in the discrete
368 space. However, this is not always possible, e.g., with spatially-varying SSMs or state-switching SSMs.
369 Following Pedersen et al. (2011), we can derive a transition rate matrix for the HMM that mimics the state
370 process of the SSM using a partial differential equation (PDE; e.g. see Okubo and Levin, 2001). There
371 are several methods (e.g., finite differencing, finite volume, or finite element) to convert this PDE into a
372 transition matrix (Quarteroni and Valli, 2008). The parameters of the SSM determine the switching rates

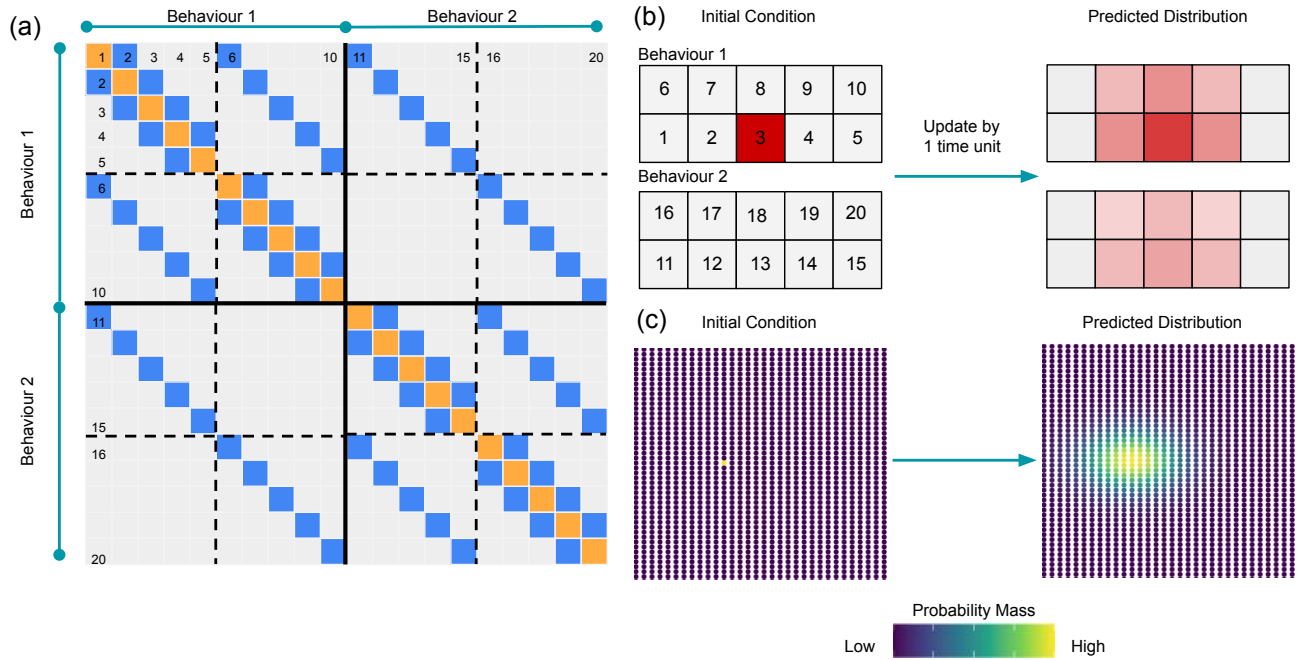


Fig. 6. Example of hidden Markov model for state-switching animal movement over 2D (x, y) space: (a) the transition rate matrix derived from the continuous-space, continuous-time partial differential equation model (Pedersen et al., 2011) where non-zero entries are coloured and some rows and columns 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-behavioural 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.

373 for the approximating HMM. Fig. 6 (Panel a) shows an example of a derived transition rate matrix for
374 a behaviour-switching Brownian motion model on the 5×2 grid. Note that for most PDE methods the
375 derived transition matrix is *sparse* (i.e. most of its entries are zero).

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

381 5.2 Pitfalls

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

394 The second pitfall is encountered when deciding whether to compute the transition probabilities directly 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
398 no thresholding is necessary; however, the PDE approach is, at present, limited to SSM with Gaussian
399 state processes and requires one to compute a sparse matrix *exponential*. Future research is needed to
400 expand one or both of these approaches to a wider range of SSMs.
401

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

412 **5.3 Opportunities**

413 This method could have wide application in ecology if future research into its pitfalls led to improved
414 computational and statistical implementation. To show this, we consider three areas where this method

415 has shown some promise, but could be further exploited: animal movement, population dynamics, and
416 encounter modelling.

417 For animal movement, our example described one possible approach (Pedersen et al., 2008; Thygesen
418 et al., 2009; Pedersen et al., 2011). It continues to be used (Jonsen et al., 2013; Braun et al., 2018; Haase et al.,
419 2021), but only in limited contexts (focusing on light-based geolocation). Despite this, the flexibility of
420 the HMM approach offers solutions to many common modelling needs that arise with animal telemetry:
421 non-Gaussian, multivariate observations are easily accommodated, unlike common alternatives based
422 on Kalman filtering (Johnson et al., 2008); continuous-time behaviour-switching can be accounted for
423 and inference on where, not just when, behaviours are exhibited is possible (Pedersen et al., 2011), while
424 alternatives make it more difficult to make this connection (see Section 6); barriers and obstacles to
425 movement (e.g., land for marine animals) are trivially accounted for, while this remains difficult for
426 simulation-based techniques; and environmental covariates can drive step-selection in continuous time,
427 similar to continuous-time discrete-space models (Hanks et al., 2015), but with flexibility to include
428 behaviour-switching and link movement to a continuous-space model.

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

438 For encounter models, we refer to applications where animal movement is inferred from encounters
439 animals have with detectors, e.g., human observers, cameras, or acoustic devices. The approximate HMM
440 is easily extended to allow for custom detection models. As with telemetry data, these detections can
441 be used to infer individual animal movement (Pedersen and Weng, 2013; Dorazio and Price, 2019). Yet,
442 more powerfully, these individual animal movement models can be incorporated into population-level
443 models such as distance sampling or spatial capture-recapture (e.g. McClintock et al., 2021). In each of
444 these methods, a critical quantity is the probability *any* given animal in the population could be detected.
445 When animals can move, this detection probability is calculated by averaging over all the possible (yet
446 unobserved) paths an individual animal *could* have travelled. This is precisely the quantity the forward
447 algorithm for HMMs can be used to compute. This approach has been taken to incorporate animal
448 movement into both distance sampling (Glennie et al., 2021) and spatial capture-recapture (Glennie et al.,
449 2019). Furthermore, this methods provides the opportunity to build a general statistical approach to
450 encounter modelling (Gurarie and Ovaskainen, 2013).

451 **6 Spatial analysis and spatial interpretation of decoded states**

452 Inferring the behavioural state of an animal based on telemetry data is an increasingly popular applic-
453 ation of HMMs in ecology. Telemetry data are primarily made up of geographical locations, which are
454 inherently spatial, and HMMs for such data often characterise animal movement as a time series of
455 steps and turns (e.g. Langrock et al., 2012), either on the plane (2D) or in a volume (3D). Inferences from
456 movement HMMs can therefore assign behavioural states to spatial locations, which can be visualised
457 in space by plotting the predicted state assignments on a map. Under these basic HMMs, the model is
458 blind to the spatial mechanisms that give rise to certain behaviours. The spatial locations of the decoded
459 states are therefore irrelevant to the model, but highly relevant to the ecological application, and the

460 ecological interpretation of decoded states. A common extension to bring spatial mechanism into the
461 model is to incorporate spatial covariate effects on the transition probabilities or parameters of the state-
462 dependent observation distributions (e.g. Morales et al., 2004; Langrock et al., 2012; McClintock et al.,
463 2012; Rivest et al., 2016; Mul et al., 2020). This creates an explicit link between space and movement
464 behaviour. The pitfall, however, with this approach is that the appropriate structure for a given research
465 question can quickly lead to a model that is very complex, has a large number of parameters, and whose
466 results are difficult to interpret. For these reasons, an attractive alternative is a two-stage approach,
467 where a relatively simple HMM is first used to get state assignments for spatial locations, and post-hoc
468 analyses are conducted to infer the relationship between the states and the spatial context in which they
469 occur (e.g. Breed et al., 2009; Nickel et al., 2021). Though this is a simpler approach, there has been little
470 research into how best to build a statistically robust and rigorous two-stage model.

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

475 **6.1 Pitfall example: Do hummingbirds use landmarks to remember spatial locations?**

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

482 The objective of the study was to gain a better understanding of how hummingbirds learn rewarding
 483 spatial locations, and as such there was an interest in *where*, with respect to the feeder and landmarks,
 484 certain movement behaviours occurred, as illustrated in Fig. 7. Knowing if birds were more likely to
 485 display targeted searching behaviour in particular parts of space would allow inferences about what
 486 specific features were learned and remembered. Hypothesis tests (t-tests) were carried out to determine
 487 whether inferred searching state locations and stops made by experienced birds (with several prior visits
 488 to the feeder in the presence of landmarks) were closer to the feeder location than would have been
 489 expected by chance, compared to naive birds with a single prior visit. While this is a reasonable approach
 490 and provides some insight into where behaviours occur, it does not fully utilise the information contained
 491 in the spatial distribution of states, or account for uncertainty in the state assignments.

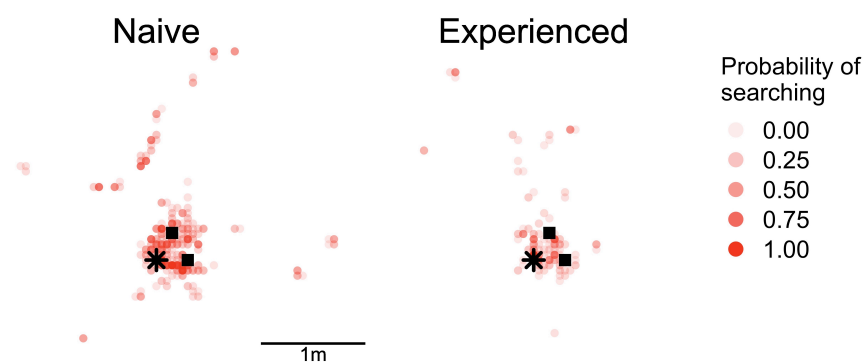


Fig. 7. The spatial distribution of locations assigned the searching state for naive birds, who only had one prior visit to the feeder (asterisk) before it was removed, and experienced birds who had several prior visits. The black squares mark the location of two artificial landmarks that were present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

492 6.2 Opportunities

493 This pitfall is at odds with what many, if not most, ecological and conservation science studies want
 494 to achieve when analysing telemetry data: an understanding of why behaviours occur where they

495 do. In addition, the most easily implemented conservation and management interventions are often
496 place-based, for example Marine Protected Areas. To inform place-based management decisions with
497 inferences from movement HMMs, we must be able to identify both important habitat (e.g. stop-over
498 sites) and the behaviour(s) associated with it (e.g. migration). This is key for identifying what parts of
499 space have disproportionate functional value to an animal population and deserve further protection
500 or management (Lennox et al., 2019). Extending the use of HMMs to better answer these questions or
501 promoting alternative methods will have widespread impact on conservation practice, feeding into policy
502 and management decisions about place- and time-based conservation of wildlife, as well as their spatial
503 and temporal exposure to risk.

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

518 using a discrete-time approximation (e.g. Equation (S1) in Section S1.1 of the Supplementary Material),
519 as has been recently done (e.g. `momentuHMM` version 2.0.0; McClintock and Michelot, 2018).

520 **7 Conclusion**

521 Hidden Markov models are a versatile class of models that will continue to be developed for and applied
522 to ecological problems. In this review, we have highlighted capabilities of HMMs that could be fruitful
523 for ecological applications, but where one must be aware of current statistical and computational pitfalls.
524 HMMs are becoming more complex in structure to better realise the potential of the more heterogeneous,
525 multi-scale, multi-dimensional data being collected and the more detailed research questions being asked.
526 This can make HMMs a valuable tool to capture a wide variety of observations, to model multiple
527 processes acting at difference scales, to describe individual and population-level effects, and to link state-
528 switching to space as well as time. With these developments, however, come greater difficulties in inter-
529 preting these models, justifying their assumptions, and fitting these models with current computational
530 capabilities. We hope this paper heightens awareness of the trouble more complex models can bring and
531 what opportunities for future solutions these problems can inspire.

532 **Acknowledgements**

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

537 **Data Availability**

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

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., Mills Flem-
558 ming, J., Nielsen, A., Petris, G., et al. (2021). A guide to state–space modeling of ecological time series.
559 *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*
567 *arXiv:1807.11907*.
- 568 Blackwell, P. G., Niu, M., Lambert, M. S., and LaPoint, S. D. (2016). Exact Bayesian inference for animal
569 movement in continuous time. *Methods in Ecology and Evolution*, 7(2):184–195.
- 570 Borchers, D., Distiller, G., Foster, R., Harmsen, B., and Milazzo, L. (2014). Continuous-time spatially
571 explicit capture–recapture models, with an application to a jaguar camera-trap survey. *Methods in*
572 *Ecology and Evolution*, 5(7):656–665.
- 573 Braun, C. D., Galuardi, B., and Thorrold, S. R. (2018). HMMoce: An R package for improved geolocation
574 of archival-tagged fishes using a hidden Markov method. *Methods in Ecology and Evolution*, 9(5):1212–
575 1220.
- 576 Breed, G. A., Jonsen, I. D., Myers, R. A., Bowen, W. D., and Leonard, M. L. (2009). Sex-specific, seasonal
577 foraging tactics of adult grey seals (*halichoerus grypus*) revealed by state–space analysis. *Ecology*,
578 90(11):3209–3221.
- 579 Bureau, A., Shiboski, S., and Hughes, J. P. (2003). Applications of continuous time hidden Markov models
580 to the study of misclassified disease outcomes. *Statistics in Medicine*, 22(3):441–462.

- 581 Choquet, R., Garnier, A., Awuve, E., and Besnard, A. (2017). Transient state estimation using continuous-
582 time processes applied to opportunistic capture–recapture data. *Ecological Modelling*, 361:157–163.
- 583 De Valpine, P. and Hastings, A. (2002). Fitting population models incorporating process noise and
584 observation error. *Ecological Monographs*, 72(1):57–76.
- 585 Descamps, S., Tarrow, A., Cherel, Y., Delord, K., Godø, O., Kato, A., Krafft, B., Lorentsen, S., Ropert-
586 Coudert, Y., Skaret, G., and Varpe, Ø. (2016a). Data from: At-sea distribution and prey selection of
587 Antarctic petrels and commercial krill fisheries.
- 588 Descamps, S., Tarrow, A., Cherel, Y., Delord, K., Godø, O. R., Kato, A., Krafft, B. A., Lorentsen, S.-H.,
589 Ropert-Coudert, Y., Skaret, G., et al. (2016b). At-sea distribution and prey selection of Antarctic petrels
590 and commercial krill fisheries. *PloS one*, 11(8):e0156968.
- 591 Dorazio, R. M. and Price, M. (2019). State-space models to infer movements and behavior of fish detected
592 in a spatial array of acoustic receivers. *Canadian Journal of Fisheries and Aquatic Sciences*, 76(4):543–
593 550.
- 594 Fine, S., Singer, Y., and Tishby, N. (1998). The hierarchical hidden Markov model: analysis and
595 applications. *Machine Learning*, 32(1):41–62.
- 596 Garcke, J. (2012). Sparse grids in a nutshell. In *Sparse grids and Applications*, pages 57–80. Springer.
- 597 Gatti, P., Fisher, J. A., Cyr, F., Galbraith, P. S., Robert, D., and Le Bris, A. (2021). A review and tests of
598 validation and sensitivity of geolocation models for marine fish tracking. *Fish and Fisheries*, 22(5):1041–
599 1066.
- 600 Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A., and Rubin, D. B. (2013). *Bayesian data*
601 *analysis*. CRC press.

- 602 Glennie, R., Borchers, D. L., Murchie, M., Harmsen, B. J., and Foster, R. J. (2019). Open population
603 maximum likelihood spatial capture-recapture. *Biometrics*, 75(4):1345–1355.
- 604 Glennie, R., Buckland, S. T., Langrock, R., Gerrodette, T., Ballance, L., Chivers, S., and Scott, M. (2021).
605 Incorporating animal movement into distance sampling. *Journal of the American Statistical Association*,
606 116(533):107–115.
- 607 Gurarie, E. and Ovaskainen, O. (2013). Towards a general formalization of encounter rates in ecology.
608 *Theoretical Ecology*, 6(2):189–202.
- 609 Haase, S., Krumme, U., Gräwe, U., Braun, C. D., and Temming, A. (2021). Validation approaches of a
610 geolocation framework to reconstruct movements of demersal fish equipped with data storage tags in
611 a stratified environment. *Fisheries Research*, 237:105884.
- 612 Hanks, E. M., Hooten, M. B., and Alldredge, M. W. (2015). Continuous-time discrete-space models for
613 animal movement. *The Annals of Applied Statistics*, 9(1):145–165.
- 614 Jackson, C. H. (2011). Multi-state models for panel data: The msm package for R. *Journal of Statistical*
615 *Software*, 38(8):1–29.
- 616 Jackson, C. H., Sharples, L. D., Thompson, S. G., Duffy, S. W., and Couto, E. (2003). Multistate Markov
617 models for disease progression with classification error. *Journal of the Royal Statistical Society: Series*
618 *D (The Statistician)*, 52(2):193–209.
- 619 Johnson, D. S., London, J. M., Lea, M.-A., and Durban, J. W. (2008). Continuous-time correlated random
620 walk model for animal telemetry data. *Ecology*, 89(5):1208–1215.

- 621 Jonsen, I., Basson, M., Bestley, S., Bravington, M., Patterson, T., Pedersen, M. W., Thomson, R., Thygesen,
622 U. H., and Wotherspoon, S. (2013). State-space models for bio-loggers: A methodological road map.
623 *Deep Sea Research Part II: Topical Studies in Oceanography*, 88:34–46.
- 624 Kays, R., Crofoot, M. C., Jetz, W., and Wikelski, M. (2015). Terrestrial animal tracking as an eye on life
625 and planet. *Science*, 348(6240).
- 626 Kitagawa, G. (1987). Non-Gaussian state–space modeling of nonstationary time series. *Journal of the*
627 *American statistical association*, 82(400):1032–1041.
- 628 Langrock, R. and King, R. (2013). Maximum likelihood estimation of mark–recapture–recovery models
629 in the presence of continuous covariates. *The Annals of Applied Statistics*, 7(3):1709–1732.
- 630 Langrock, R., King, R., Matthiopoulos, J., Thomas, L., Fortin, D., and Morales, J. M. (2012). Flexible
631 and practical modeling of animal telemetry data: hidden Markov models and extensions. *Ecology*,
632 93(11):2336–2342.
- 633 Langrock, R. and Zucchini, W. (2011). Hidden markov models with arbitrary state dwell-time distribu-
634 tions. *Computational Statistics & Data Analysis*, 55(1):715–724.
- 635 Lennox, R. J., Aarestrup, K., Cooke, S. J., Cowley, P. D., Deng, Z. D., Fisk, A. T., Harcourt, R. G., Heupel, M.,
636 Hinch, S. G., Holland, K. N., et al. (2017). Envisioning the future of aquatic animal tracking: technology,
637 science, and application. *BioScience*, 67(10):884–896.
- 638 Lennox, R. J., Engler-Palma, C., Kowarski, K., Filous, A., Whitlock, R., Cooke, S. J., and Auger-Méthé, M.
639 (2019). Optimizing marine spatial plans with animal tracking data. *Canadian Journal of Fisheries and*
640 *Aquatic Sciences*, 76(3):497–509.

- 641 Leos-Barajas, V., Gangloff, E. J., Adam, T., Langrock, R., Van Beest, F. M., Nabe-Nielsen, J., and Morales,
642 J. M. (2017). Multi-scale modeling of animal movement and general behavior data using hidden Markov
643 models with hierarchical structures. *Journal of Agricultural, Biological and Environmental Statistics*,
644 22(3):232–248.
- 645 Liu, Y.-Y., Li, S., Li, F., Song, L., and Rehg, J. M. (2015). Efficient learning of continuous-time hidden
646 Markov models for disease progression. *Advances in Neural Information Processing Systems*, 28:3500–
647 3607.
- 648 MacKenzie, D. I., Nichols, J. D., Royle, J. A., Pollock, K. H., Bailey, L., and Hines, J. E. (2018). *Occupancy*
649 *Estimation and Modeling: Inferring Patterns and Dynamics of Species Occurrence*. Elsevier, 2nd edition.
- 650 McClintock, B. T. (2017). Incorporating telemetry error into hidden Markov models of animal movement
651 using multiple imputation. *Journal of Agricultural, Biological and Environmental Statistics*, 22(3):249–
652 269.
- 653 McClintock, B. T. (2021). Worth the effort? A practical examination of random effects in hidden Markov
654 models for animal telemetry data. *Methods in Ecology and Evolution*, 12(8):1475–1497.
- 655 McClintock, B. T., Abrahms, B., Chandler, R. B., Conn, P. B., Converse, S. J., Emmet, R., Gardner, B.,
656 Hostetter, N. J., and Johnson, D. S. (2021). An integrated path for spatial capture-recapture and animal
657 movement modeling. *Ecology*, Accepted Author Manuscript e03473. <https://doi.org/10.1002/ecy.3473>.
- 658 McClintock, B. T., Johnson, D. S., Hooten, M. B., Ver Hoef, J. M., and Morales, J. M. (2014). When to be
659 discrete: the importance of time formulation in understanding animal movement. *Movement Ecology*,
660 2(1):1–14.

- 661 McClintock, B. T., King, R., Thomas, L., Matthiopoulos, J., McConnell, B. J., and Morales, J. M. (2012).
662 A general discrete-time modeling framework for animal movement using multistate random walks.
663 *Ecological Monographs*, 82(3):335–349.
- 664 McClintock, B. T., Langrock, R., Gimenez, O., Cam, E., Borchers, D. L., Glennie, R., and Patterson, T. A.
665 (2020). Uncovering ecological state dynamics with hidden Markov models. *Ecology Letters*, 23(12):1878–
666 1903.
- 667 McClintock, B. T. and Michelot, T. (2018). momentuHMM: R package for generalized hidden Markov
668 models of animal movement. *Methods in Ecology and Evolution*, 9(6):1518–1530.
- 669 McLachlan, G. J. and Krishnan, T. (2007). *The EM algorithm and extensions*. John Wiley & Sons.
- 670 Mews, S., Langrock, R., King, R., and Quick, N. (2020a). Continuous-time multi-state capture-recapture
671 models. *arXiv preprint arXiv:2002.10997*.
- 672 Mews, S., Langrock, R., Ötting, M., Yaqine, H., and Reinecke, J. (2020b). Maximum approximate likelihood
673 estimation of general continuous-time state-space models. *arXiv preprint arXiv:2010.14883*.
- 674 Michelot, T. and Blackwell, P. G. (2019). State-switching continuous-time correlated random walks.
675 *Methods in Ecology and Evolution*, 10(5):637–649.
- 676 Michelot, T., Gloaguen, P., Blackwell, P. G., and Étienne, M.-P. (2019). The Langevin diffusion as a
677 continuous-time model of animal movement and habitat selection. *Methods in Ecology and Evolution*,
678 10(11):1894–1907.
- 679 Morales, J. M., Haydon, D. T., Frair, J., Holsinger, K. E., and Fryxell, J. M. (2004). Extracting more out of
680 relocation data: Building movement models as mixtures of random walks. *Ecology*, 85(9):2436–2445.

- 681 Mul, E., Blanchet, M.-A., McClintock, B. T., Grecian, W. J., Biuw, M., and Rikardsen, A. (2020). Killer
682 whales are attracted to herring fishing vessels. *Marine Ecology Progress Series*, 652:1–13.
- 683 Nickel, B. A., Suraci, J. P., Nisi, A. C., and Wilmers, C. C. (2021). Energetics and fear of humans constrain
684 the spatial ecology of pumas. *Proceedings of the National Academy of Sciences*, 118(5):e2004592118.
- 685 Okubo, A. and Levin, S. A. (2001). *Diffusion and ecological problems: modern perspectives*. Springer.
- 686 Patterson, T. A., Parton, A., Langrock, R., Blackwell, P. G., Thomas, L., and King, R. (2017). Statistical
687 modelling of individual animal movement: an overview of key methods and a discussion of practical
688 challenges. *AStA Advances in Statistical Analysis*, 101(4):399–438.
- 689 Pedersen, M. W., Patterson, T. A., Thygesen, U. H., and Madsen, H. (2011). Estimating animal behavior
690 and residency from movement data. *Oikos*, 120(9):1281–1290.
- 691 Pedersen, M. W., Righton, D., Thygesen, U. H., Andersen, K. H., and Madsen, H. (2008). Geolocation of
692 North Sea cod (*Gadus morhua*) using hidden markov models and behavioural switching. *Canadian
693 Journal of Fisheries and Aquatic Sciences*, 65(11):2367–2377.
- 694 Pedersen, M. W. and Weng, K. C. (2013). Estimating individual animal movement from observation
695 networks. *Methods in Ecology and Evolution*, 4(10):920–929.
- 696 Plummer, M. (2017). *JAGS: Just Another Gibbs Sampler*. Version 4.0.3 User Manual.
- 697 Pohle, J., Langrock, R., van Beest, F. M., and Schmidt, N. M. (2017). Selecting the number of states in hid-
698 den Markov models: pragmatic solutions illustrated using animal movement. *Journal of Agricultural,
699 Biological and Environmental Statistics*, 22(3):270–293.
- 700 Pritchard, D. J., Hurly, T. A., Photopoulou, T., and Healy, S. D. (2021). 3d flightpaths reveal the
701 development of spatial memory in wild hummingbirds. *In prep*.

- 702 Quarteroni, A. and Valli, A. (2008). *Numerical approximation of partial differential equations*. Springer
703 Science & Business Media.
- 704 R Core Team (2020). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical
705 Computing, Vienna, Austria.
- 706 Rivest, L.-P., Duchesne, T., Nicosia, A., and Fortin, D. (2016). A general angular regression model for
707 the analysis of data on animal movement in ecology. *Journal of the Royal Statistical Society: Series C*
708 (*Applied Statistics*), 65(3):445–463.
- 709 Royle, J. A. and Kéry, M. (2007). A Bayesian state-space formulation of dynamic occupancy models.
710 *Ecology*, 88(7):1813–1823.
- 711 Rubin, D. B. (2004). *Multiple imputation for nonresponse in surveys*. John Wiley & Sons.
- 712 Sacchi, G. and Swallow, B. (2021). Toward efficient Bayesian approaches to inference in hierarchical
713 hidden markov models for inferring animal behavior. *Frontiers in Ecology and Evolution*, 9:249.
- 714 Sherlock, C. (2021). Direct statistical inference for finite markov jump processes via the matrix exponen-
715 tial. *Computational Statistics*.
- 716 Sidje, R. B. (1998). Expokit: A software package for computing matrix exponentials. *ACM Transactions*
717 *on Mathematical Software (TOMS)*, 24(1):130–156.
- 718 Sidrow, E., Heckman, N., Fortune, S. M., Trites, A. W., Murphy, I., and Auger-Méthé, M. (2021).
719 Modelling multi-scale state-switching functional data with hidden markov models. *arXiv preprint*
720 *arXiv:2101.03268*.

- 721 Thygesen, U. H., Pedersen, M. W., and Madsen, H. (2009). Geolocating fish using hidden Markov models
722 and data storage tags. In *Tagging and Tracking of Marine Animals with Electronic Devices*, pages 277–293.
723 Springer.
- 724 Zucchini, W., MacDonald, I. L., and Langrock, R. (2017). *Hidden Markov models for time series: an*
725 *introduction using R, Second Edition*. CRC press.