RESEARCH

Incorporating Periodic Variability in Hidden Markov Models for Animal Movement

Michael Li[1]* and Benjamin M. Bolker[1,2]

Sample of title note

Correspondence:
m88@mcmaster.ca
Department of Biology,
McMaster University, 1280 Main
t. West, L8S 4K1, Hamilton,
Ontario, Canada
ull list of author information is
vailable at the end of the article
Equal contributor

Abstract

Background: Clustering time-series data into discrete groups can improve
prediction and provide insight into the nature of underlying, unobservable states
of the system. However, temporal variation in the rates at which individuals move
between groups can obscure such signals. We use finite mixture and hidden
Markov models (HMMs), two standard clustering techniques, to model
high-resolution hourly movement data from Florida panthers (*Puma concolor
coryi*). Allowing for temporal heterogeneity in transition probabilities, a
straightforward but little-used extension of the standard HMM framework,
resolves some shortcomings of current models and clarifies the behavioural
patterns of panthers.

Results: Simulations and Florida panthers data showed model misspecification
(omitting important sources of variation) can lead to overfitting and
over-estimating number of behavioural states. Models incorporating temporal
heterogeneity have lower number of states with slightly higher variation in short
movement states, and able to make out of sample predictions that captures
observed diurnal and autocorrelation patterns exhibited by Florida panthers.

Conclusion: Incorportating temporal heterogeneity reduce the selected number
of behavioural states closer to a biologically interpretable level, improved
goodness of fit and predictability. Our suggest that incorporating previously
neglected structure in statistical models can allow more accurate assessment of
appropriate model complexity.

Keywords: Hidden Markov Model; Animal Movement; Temporal
Autocorrelation; Temporal Heterogeneity; Florida Panther

Background

Given a sequence of animal movements, movement models aim to find a parsimo-
nious description that can be used to understand past movements and predict future
movements. Ecologists have long considered the effects of individual-level covariates

8 (sex, age, nutritional status) and environmental covariates (habitat type, location

9 of predators or prey) on movement [1, 2, 3]. More recently, modelers have devel-

10 oped *hidden Markov models* (HMMs) [4, 5, 6] — used in animal ecology under the

11 rubric of the "multiphasic movement framework" [7] — that consider the effects of

12 organisms' *internal* states; in particular, HMMs model animal movement as though

13 individual animals' movement behaviour at particular times is determined by which

14 of a discrete set of unobserved movement states (e.g. "foraging", "traveling", "rest-

15 ing") they currently occupy. Conditional on the state occupied by an individual,

16 HMMs typically assume that animals follow a standard correlated walk model [8, 9].

17 Ever-increasing capabilities of remote sensors are making movement data avail-

18 able over an ever-wider range of time scales, at both higher resolution (e.g. hourly

19 data from GPS collars vs. daily or weekly fixes for radio or VHF collars) and longer

20 extent (e.g. from a few days to significant fractions of a year, or longer). When

21 analyzing such long-term data, ecologists will more often have to account for tem-

22 poral variability in movement behaviour at diurnal and seasonal scales that were

23 previously not captured in the data.

24 HMMs have typically been used to model movements over short time scales, where

25 the probability of transitioning between movement states is approximately constant.

26 Changes in latent/hidden behavioural state/mode transition probabilities based on

27 the local environment can be accounted for incorporating environmental covariates

28 in the HMM [10], or by more direct comparisons between inferred states and en-

29 vironmental conditions [7]. Schliehe-Diecks et al. [11] consider temporal trends in

30 behavioural transitions over the time scales of a six-hour observation period, but in

31 general ecologists have turned to other tools to describe behavioural changes over

32 longer (diurnal, seasonal, or ontogenetic) time scales [12].

33 For movement behaviours that change on a fast time scale, such that movement

34 behaviours recorded at successive observations are effectively independent, *finite*

35   *mixture models* (FMMs) — which can be considered a special case of HMMs where

36   the probability of state occupancy is independent of the previous state — can

37   adequately describe movement [13]. When movement varies over long time scales

38   (relative to the time between observations) with little short-term persistence or

39   correlation, movement could be well represented by FMMs where the occupancy

40   probabilities change deterministically over time. Thus FMMs and HMMs, with or

41   without temporal variation in the occupancy or transition probabilities, form a

42   useful family of models for capturing changes in movement behaviour over a range

43   of time scales.

44   Our primary goal in this paper is to introduce the use of HMMs with temporally

45   varying transition probabilities – in particular, transition probabilities that follow

46   a diurnal cycle – for modeling animal movement recorded over long time scales. In

47   addition to simulation-based examples, we also re-analyze data from van de Kerk et

48   al.[14], who used temporally homogeneous hidden semi-Markov models (HSMMs:

49   an extension of HMMs that allow flexible modelling of the distribution of *dwell*

50   *times*, the lengths of consecutive occupancy of a behavioural state) to describe the

51   movement and putative underlying behavioural states of Florida panthers (*Puma*

52   *concolor coryi*).

53   van de Kerk et al.[14] found that the best-fitting HSMMs incorporated a surpris-

54   ingly large number of hidden behavioural states (as many as six for individuals with

55   a large amount of available data); for reasons of computational practicality and bi-

56   ological interpretability, they restricted their detailed analysis to models with only

57   three underlying states. In contrast, most studies using HMM have chosen the num-

58   ber of underlying states *a priori*, typically using either two [11, 15, 6, 7], or three

59   states [16, 17, 18]. In contrast, [19] evaluated models with up to 10 states, but like

60   [14] they chose to consider only models with three states. As van de Kerk et al.

61   [14] comment, and as we discuss further below, behavioural repertoires with more

than three distinct states are difficult to interpret — one possible reason that other authors have not adopted van de Kerk et al.'s model-based approach to identifying the number of latent states.

Our second goal, therefore, is to explore whether van de Kerk et al.'s results on optimal model complexity might be driven at least in part by structural problems with their statistical model, i.e. the assumption of temporally homogeneous behaviour. For large data sets, information-theoretic model selection methods will typically choose complex, highly parameterized models; when there is only one way in which models can become more complex (e.g. by increasing the number of latent states), complexity that is present in the data but not accounted for in the model (e.g. spatial or temporal heterogeneity) can be misidentified as other forms of complexity. We predict that increasing volumes of data will increasingly lead researchers who are accustomed to fitting small models to sparse data into such traps. We examine whether allowing for diurnal variation in the Florida panther data leads to selection of models with smaller numbers of latent states; we also fit models to simulated data with varying numbers of latent states and degrees of temporal heterogeneity to test our conjecture that heterogeneity can be misidentified as behavioural complexity.

Methods

Data and previous analyses

GPS collars were fitted to 18 Florida panthers in 2005-2012 by Florida Fish and Wildlife and Conservation Commission staff using trained hounds and houndsmen. Of these animals, 13 had sufficient data to be used by van de Kerk et al.[14]. Here we focus on the four cats with the most data (all with approximately 10,000-15,000 observations: see Table 1 in Supplementary Material), in part because our goal is to understand the issues that arise when simple models are fitted to large data sets, and in part because the general trend in telemetry studies is toward larger data sets. As is typical in studies of animal movement, we took first differences of the data by

decomposing contiguous sequences of hourly GPS coordinates into successive step

lengths (in meters) and turning angles (in radians) [9, 14].

van de Kerk et al.[14] used hidden semi-Markov models (HSMM), an extension of

HMM that permits explicit modelling of dwell times [6], considering both Poisson

and negative binomial distributions for dwell times. As shown by van de Kerk et

al.[14] (Figure S3b, top row, middle panel), the estimated shape parameter of the

negative binomial dwell time distribution was typically close to 1 ($\approx 0.4 - 1.6$;

confidence intervals were not given), implying that a geometric distribution (i.e.,

negative binomial with shape=1) might be adequate. In turn, this suggests that we

might not lose much accuracy by reverting to a simpler HMM framework, which

corresponds to making precisely this assumption.

van de Kerk et al.[14] considered time-homogeneous models with a variety of

candidate distributions — log-Normal, Gamma, and Weibull distributions for step

lengths and von Mises and wrapped Cauchy distributions for the turning angle

— concluding on the basis of the Akaike information criterion (AIC) that Weibull

step length and wrapped Cauchy turning angle distributions were best. Since our

analysis aims for simplicity and qualitative conclusions rather than for picking the

very best predictive model, we focus on models that treat each step as a univariate,

log-Normally distributed observation, glossing over both the differences in shape

between the three candidate step-length distributions and the effects of consider-

ing multivariate (i.e., step length plus turning angle) observations. However, we do

briefly compare log-Normal and Weibull step-length distributions, with and without

a von Mises-distributed turning angle included in the model (Figure 2). (Note that

most movement analyses, including van de Kerk et al. [14], are only partially multi-

variate, treating step length and turning angle at a particular time as multivariate

observations for the purpose of HMM analysis but neglecting possible correlations

between the two measures.)

₁₁₆ van de Kerk et al.[14] used the Bayesian (Schwarz) information criterion (BIC)

₁₁₇ to test the relative penalized goodness of fit for models ranging from 2 to 6 latent

₁₁₈ states. In general, BIC values decreased as the number of states increased from

₁₁₉ three to six states, suggesting that the six-state model was favoured statistically;

₁₂₀ however, the authors used three-state models in most of their analyses for ease of

₁₂₁ biological interpretation. We follow van de Kerk et al.[14] in using BIC-optimality

₁₂₂ (i.e., minimum BIC across a family of models) as the criterion for identifying the

₁₂₃ best model, because we are interested in explaining the data generation process by

₁₂₄ identifying the "true" number of underlying movement states.

₁₂₅ Using BIC also simplifies evaluation of model selection procedures; it is easier to

₁₂₆ test whether our model selection procedure has selected the model used to simulate

₁₂₇ the data, rather than testing whether it has selected the model with the minimal

₁₂₈ Kullback-Leibler distance [20]. We recognize that ecologists will often be interested

₁₂₉ in maximizing predictive accuracy rather than selecting a true model, and that as

₁₃₀ usual in ecological systems the true model will be far more complicated than any

₁₃₁ candidate model [21]; we believe that the qualitative conclusions stated here for

₁₃₂ BIC-optimality will carry over to analyses using AIC instead.

₁₃₃ Model description

₁₃₄ In a HMM, the joint likelihood of *emissions* (i.e., direct observations) $\mathbf{Y} = \mathbf{y}_1, ..., \mathbf{y}_T$

₁₃₅ and a hidden state sequence $\mathbf{Z}, z_t \in \{1, ..., n\}, t = 1, ..., T$, given model parameters

₁₃₆ $\boldsymbol{\theta}$ and covariates $\mathbf{X}_{1:T} = \mathbf{x}_1, ..., \mathbf{x}_T$, can be written as:

$$P(\mathbf{Y}_{1:T}, \mathbf{Z}_{1:T}|\boldsymbol{\theta}, \mathbf{X}_{1:T}) = P(z_1 \mid \mathbf{x}_1)P(\mathbf{y}_1|z_1, \mathbf{x}_1) \cdot$$
$$\prod_{k=2}^{T} P(z_k|z_{k-1}, \mathbf{x}_k)P(\mathbf{y}_k|z_k, \mathbf{x}_k) \tag{1}$$

137   The emissions $\mathbf{y}_i$ are boldfaced to denote that we may have a vector of observations

138   at each time point (e.g., step length and turning angle). The model contains three

139   distinct components:

140   **Initial probability** $P(z_1 = i|\mathbf{x_1})P(\mathbf{y_1}|z_1, \mathbf{x_1})$: the probability of state i at time

141        $t = 1$ where the covariate is $\mathbf{x_1}$, times the vector of observations $\mathbf{y_1}$ condi-

142   tioned on covariates $\mathbf{x_1}$ and state z_1.

143   **Transition probability** $P(z_k = j|z_{k-1} = i, \mathbf{x_k})$: the probability of a transition

144   from state i at time $t = k - 1$ to state j with covariate $\mathbf{x_k}$ at time $t = k$.

145   **Emission probability** $P(\mathbf{y_k}|z_k, \mathbf{x_k})$: a vector of observations $\mathbf{y_k}$ conditioned on

146   covariates $\mathbf{x_k}$ at state z_k at time $t = k$.

147   Eq. 1 gives the likelihood of the observed sequence given (conditional on) a partic-

148   ular hidden sequence. In order to calculate the overall, unconditional (or marginal)

149   likelihood of the observed sequence, we need to average over all possible hidden

150   sequences. There are several efficient algorithms for computing the marginal like-

151   lihood and numerically estimating parameters [22]; we used those implemented in

152   the `depmixS4` package for R [23, 24].

153   For any n-state HMM, we need to define a $n \times n$ matrix that specifies the

154   probabilities π_{ij} of being in movement states j at time $t+1$ given that the individual

155   is in state i. The FMM is a special case of HMM where the probabilities of *entering*

156   a given state are identical across all states — i.e., the probability of occupying a

157   state at the next time step is independent of the current state occupancy. It can be

158   modelled in the HMM framework by setting the transition probabilities $\pi_{ij} = \pi_{i*}$.

159   In any case, the transition matrix π_{ij} must respect the constraints that (1) all

160   probabilities are between 0 and 1 and (2) transition probabilities out of a given state

161   sum to 1. As is standard for HMMs with covariates [23], we define this multinomial

162   logistic model in terms of a linear predictor η_{ij}, where η_{i1} is set to 1 without loss

163   of generality (i.e. we have only $n \times (n-1)$ distinct parameters; we index j from

164   2 to n for notational clarity):

$$\pi_{ij} = \exp(\eta_{ij}(t)) / \left(1 + \sum_{j=2}^{n} \exp(\eta_{ij}(t)) \right), \text{ for } j = 2, ..., n$$

$$\pi_{i1} = 1 - \sum_{j=2}^{n} \pi_{ij}$$

(2)

165   We considered four different transition models for diurnal variation in behaviour,

166   incorporating hour-of-day as a covariate following the general approach of Morales

167   et al.[17] of incorporating covariate dependence in the transition matrix.

168   **Multiple block transition** Here we assume piecewise-constant transition prob-

169   abilities. The transition probability π_{ij} is a function of time (hour of day),

170   where it is assigned to one of M different time blocks:

$$\eta_{ij}(t) = \sum_{m=1}^{M} a_{ijm} \delta_{m=t}$$

171   where a_{ijm} are parameters, and $\delta_{m=t}$ is a Kronecker delta ($\delta_{m=t} = 1$ for

172   the time block at the corresponding time t, and 0 otherwise).

173   **Quadratic transition model** We assume the elements of the linear predictor are

174   quadratic functions of hour. The quadratic model is not diurnally continuous,

175   i.e. there is no constraint that forces $\eta_{ij}(0) = \eta_{ij}(24)$; imposing a diurnal

176   continuity constraint would collapse the model to a constant.

$$\eta_{ij}(t) = b_{ij1} + b_{ij2} \left(\frac{t}{24} \right) + b_{ij3} \left(\frac{t}{24} \right)^2$$

177 **Sinusoidal transition model** A sinusoidal model with a period of 24 hours is

178   identical in complexity to the quadratic model, but automatically satisfies

179   the diurnal continuity constraint.

$$\eta_{ij}(t) = b_{ij1} + b_{ij2}\cos\left(\frac{2\pi t}{24}\right) + b_{ij3}\sin\left(\frac{2\pi t}{24}\right)$$

180 **Hourly model** Lastly, we extended the multi-block approach and assign a differ-

181   ent transition matrix for every hour of the day. This model is included for

182   comparative purposes due to the large number of parameters in the model

183   which makes it not really practical. We only fitted up to four states using the

184   hourly model.

185   Other periodic functions, such as Fourier series (the sinusoidal transition model

186 augmented by additional sinusoidal components at higher frequencies) or periodic

187 splines, could be useful directions for future exploration.

188   Model evaluation

189 We used the `depmixS4` package to fit covariate-dependent transition HMMs, simu-

190 late states and step lengths using the estimated parameters, and estimate the most

191 likely states with the Viterbi algorithm.

192   We used three approaches to assess the fit of both time-homogeneous and time-

193 inhomogeneous HMMs with 3 to 6 states to step-length data from the four of the

194 thirteen Florida panthers with the most data ($>$ **9000** observations). (1) Comparing

195 BICs to the optimal-BIC model within each type of transition complexity ($\mathbf{\Delta}$BIC

196 = BIC - min(BIC)) assesses the overall goodness of fit of each model type. (2)

197 Comparing average step-length by hour of day for the observed data and for data

198 simulated from the models shows how well a particular class of models can capture

199 the diurnal variation in behaviour. (3) Comparing temporal autocorrelations for the

200   observed data and for data simulated from the models shows how well a particular

201   class of models captures serial correlation at both short and long scales.

202   Model complexity and the number of parameters increase as the number of latent

203   states increase, FMM to HMM, and lastly, FMM and HMM incorporating temporal

204   heterogeneity. The number of free parameters in an HMM can be generalized by

205   summing up the number of free parameters of the three distinct components. Let

206   n be the number of hidden states and k_i, k_t, k_e be the number of parameters

207   describing the covariate-dependence of the prior distribution, transition function

208   and emission distributions; that is, for a homogeneous model, $k = 1$, while a single

209   numeric covariate or a categorical predictor with two levels would give $k = 2$. Then

210   the number of free parameters of an HMM is:

$$\text{Number of Free Parameter} = \underbrace{k_i \cdot (n-1)}_{\text{Initial}} + \underbrace{k_t \cdot n \cdot (n-1)}_{\text{Transition}} + \underbrace{k_e \cdot n}_{\text{Emission}} \quad (3)$$

211   As the number of states increases, the number of free parameters in time-

212   homogeneous FMMs and HMMs and FMMs with temporal heterogenity will in-

213   crease linearly, whereas HMMs with temporal heterogenity will increase quadrat-

214   ically (Eq. 3). When comparing BICs, it is important to account for the tradeoff

215   between log-likelihood and number of states, but also log-likelihood and number of

216   free parameters.

217   We used simulations to predict hourly step length and ACF because, while the

218   computation is reasonably straightforward for FMMs, and manageable for homoge-

219   neous HMMs, the interaction between the geometric dwell time within each state

220   and the temporally varying interaction probabilities makes it unreasonably complex.

221   We used this approach to validate our models and comparing these models with the

222   observed movements instead of the standard Viterbi predictions by the Viterbi

algorithm because Viterbi predictions, which use the most probable sequence of movement states based on the observations [22, 6], double-count the observed data. It is useful to predict missing data in the observation sequence, but because it is conditional on the observed values, it can not reliably evaluate goodness of fit for the different structural complexities of HMM models.

Results

We simulated a two-state HMM with sinusoidal temporal transitions 100 times and fitted it with two to five state HMMs and without temporal transition. Heterogeneous transition models can always predict the correct number of states, whereas, can overestimate the number of states via BIC-optimal approach (Figure 1).

The BIC-optimal number of states for time homogenous models is consistent with van de Kerk et al.'s [14] results (Weibull wrapped-Cauchy to Weibull von Mises, and Weibull von Mises to log Normal without turning angles; Figure 2)

As a complement, we also fitted FMM and FMM with sinusoidal variation in state occupancy probabilities to compare the temporal effects in goodness of fit (dashed lines). As a reminder, FMMs assume that the latent state in each time step is *independent* of the latent state at the previous time step; time-varying FMMs can accurately describe movement when behaviour can change on a short time scale, but the average propensity for different behaviours changes over time.

Models with temporal heterogenity are better (lower BIC) than homogeneous models in both FMM and HMM frameworks, but time-homogenous HMMs are better than FMMs with sinusoidal temporal heterogeneity (Figure 3). Turning to the temporally heterogeneous HMMs (right panel), we see that the model with different transition probabilities for each hour of the day (HMM + THhourly) is overparameterized; it underperforms homogeneous HMM with even 3 states, and gets much worse with 4 states. The multiple-block model approximately matches the homogeneous HMM, although it gives the BIC-optimal number of states as 4, in contrast

to 6 for the homogeneous HMM. Finally, the quadratic and sinusoidal models are considerably better than any other models tested; they both give the BIC-optimal number of states as 5, and they have similar goodness of fit. However, this similarity is somewhat overstated due to the very large variation in BIC (over thousands of units) across the full range of models; there is a difference of approximately 80 BIC units, which would normally be interpreted as an enormous difference in goodness of fit, between the sinusoidal and quadratic models (both of which have 90 parameters).

The panthers exhibits a clear diurnal pattern from the average hourly step lengths from the observed data (Figure 4). As expected, temporally homogeneous models (whether FMM or HMM) predict the same mean step length regardless of time of day, failing to capture the diurnal activity cycle. All of the models incorporating temporal heterogeneity, including the temporally heterogeneous FMM, can capture the observed patterns. However, the block model does markedly worse than the other temporal models (changing the block definitions might help), and the (over-parameterized) hourly model does better than any other model at capturing the early-evening peak (but worse at capturing the mid-day trough). We also included average hourly step lengths from three-state temporally homogeneous HMM Viterbi prediction (v points).

Like the diurnal pattern (Figure 4), the strong autocorrelation of the observed step lengths at a 24-hour lag (Figure 5) shows the need to incorporate temporal heterogeneity in the model — we could have reached this conclusion even without developing any of the temporal-heterogeneity machinery. Because there are a huge number of potential complexities that can be added to movement models (e.g. spatial/temporal/among-individual heterogeneity; effects of conspecific attraction or avoidance; memory or cognitive effects), each with associated costs in researcher and computational effort, such diagnostic plots are invaluable. In contrast to the

277 hourly averages, the autocorrelation (ACF) captures both short- and long-term

278 temporal effects. HMM without temporal heterogeneity captures the short-term

279 autocorrelation, but misses the long-term autocorrelation beyond a 7-hour lag.

280 Temporally homogeneous FMM, by definition, produces neither short- nor long-

281 term autocorrelation. FMM without temporal heterogeneity, although it captures

282 the diurnal pattern well, underpredicts the degree of short-term autocorrelation.

283 The hardest problem with multiple latent states is interpreting them biologically.

284 We have no way of knowing what panthers are actually thinking (it is certainly more

285 complex than being in one of a small number of discrete latent states); we don't

286 know the "true" number of latent states, nor are we able to observe them directly,

287 although incorporating additional direct observations of behaviour (if available)

288 can at least partially address this problem [7]. Three distinct movement states seem

289 biologically interpretable for Florida panthers according to van de Kerk et al.[14]:

290 Short step length suggests resting states, intermediate step length a foraging state,

291 and long step length a traveling state.

292 The estimated parameter values for several cats (mean and standard devia-

293 tion of the step length in each state) between the time-homogeneous and time-

294 heterogeneous models are similar across all cats (Figure 7). In general, the states

295 with longer mean step lengths are relatively similar between model classes. For cats

296 14 and 15, the states with the longest or next-longest mean step lengths have similar

297 means and standard deviations; for cats 1 and 2, three long-step states in the ho-

298 mogeneous HMM appear to divide two long-step states in the heterogeneous HMM.

299 For short-step states, the heterogeneous HMM tends to identify a high-variance

300 state, while the homogeneous HMM picks up states with very short step lengths

301 (questionable in any case because we have not taken any special efforts to account

302 for GPS error).

Discussion

HMMs are a widely used and flexible tool for modeling animal movement behaviour; we need to work harder to make sure they are both appropriately complex and biologically interpretable. With the increasing volumes of movement data available, ecologists who naively use traditional homogeneous HMMs and standard information-theoretic criteria to estimate the number of behavioural states will generally overfit their data, in the sense of "discovering" large number of states that are difficult to interpret biologically.

On a broad spectrum, it really depends on what kind of question that is being answered. On one side of the spectrum, if the goal is to identify states, it might be sufficient to use a simple/traditional HMM model and pre-specify the number of states and, post hoc, match Viterbi-based states estimates with environmental variation [7]. On the other side of the spectrum, if the goal of interest is to make predictions (out of sample), it might be better to fit a covariate-dependent model so that we can explicitly model the switching process. In that case, fitting a covariate-dependent model is better for out of sample prediction because Viterbi can only estimate state occupancy if observed movements are available (within sample predictions). Finally, if we want to estimate the number of states, BIC is not necessarily good for estimation of number of states [25], but it can be useful as an approximate upper limit estimate.

Incorporating temporal heterogenity in animal movement is one step in the right direction, but much remains to be done. Our model neglects other predictors, such as habitat type or location with respect to environmental features such as roads, that can potentially improve goodness of fit and predictions and further reduce the estimated number of states. While adding more covariates is in principle straightforward using existing frameworks, including all possible biological complexities in a HMM with state-dependent transitions may rapidly become intractable in terms

of both computational time and complexity of choosing among possible reduced models and numbers of states. Better diagnostic procedures and tests are needed: these can both test overall goodness-of-fit [26] and, more importantly, localize fitting problems to particular aspects of the data so that models can be constructed without needing to include all possible features of interest.

Conclusion

We have presented a relatively simple but little-used extension (time-dependent transitions) that partly resolves the problem. Time-dependent transitions appear to offer a simple way to (1) reduce the selected number of states closer to a biologically interpretable level; (2) capture observed diurnal and autocorrelation patterns in a predictive model; (3) improve overall model fit (i.e., lower BIC) and reduce the level of complexity (number of parameters) of the most parsimonious models. Simple simulations where the true number of states is known, and transitions among states vary over time, confirm that using BIC with homogeneous HMMs overestimates the number of behavioural states, while time-dependent HMMs correctly estimate the number.

Declarations

1 Data accessibility

Hourly step lengths and turning angles of male and female Florida panthers available at:

http://ufdc.ufl.edu//IR00004241/00001.

Author's contributions

Equally contributed.

Competing interests

The authors declare that they have no competing interests.

358 **Author details**

359 [1] Department of Biology, McMaster University, 1280 Main St. West, L8S 4K1, Hamilton, Ontario, Canada. [2]

360 Department of Mathematics and Statistics, McMaster University, 1280 Main St. West, L8S 4K1, Hamilton, Ontario,

361 Canada.

362 **References**

363 1. Patterson, T.A., Thomas, L., Wilcox, C., Ovaskainen, O., Matthiopoulos, J.: State–space models of individual

364   animal movement. Trends in Ecology & Evolution **23**(2), 87–94 (2008)

365 2. McKenzie, H.W., Lewis, M.A., Merrill, E.H.: First passage time analysis of animal movement and insights into

366   the functional response. Bulletin of Mathematical Biology **71**(1), 107–129 (2009)

367 3. Pal, S., Ghosh, B., Roy, S.: Dispersal behaviour of free-ranging dogs (*Canis familiaris*) in relation to age, sex,

368   season and dispersal distance. Applied Animal Behaviour Science **61**(2), 123–132 (1998)

369 4. Firle, S., Bommarco, R., Ekbom, B., Natiello, M.: The influence of movement and resting behavior on the

370   range of three carabid beetles. Ecology **79**(6), 2113–2122 (1998).

371   doi:10.1890/0012-9658(1998)079[2113:TIOMAR]2.0.CO;2. Accessed 2015-04-14

372 5. Nathan, R., Getz, W.M., Revilla, E., Holyoak, M., Kadmon, R., Saltz, D., Smouse, P.E.: A movement ecology

373   paradigm for unifying organismal movement research. Proceedings of the National Academy of Sciences

374   **105**(49), 19052–19059 (2008). doi:10.1073/pnas.0800375105. Accessed 2015-04-29

375 6. Langrock, R., King, R., Matthiopoulos, J., Thomas, L., Fortin, D., Morales, J.M.: Flexible and practical

376   modeling of animal telemetry data: hidden Markov models and extensions. Ecology **93**(11), 2336–2342 (2012).

377   doi:10.1890/11-2241.1. Accessed 2013-10-24

378 7. Fryxell, J.M., Hazell, M., Börger, L., Dalziel, B.D., Haydon, D.T., Morales, J.M., McIntosh, T., Rosatte, R.C.:

379   Multiple movement modes by large herbivores at multiple spatiotemporal scales. Proceedings of the National

380   Academy of Sciences **105**(49), 19114–19119 (2008). doi:10.1073/pnas.0801737105. Accessed 2013-04-09

381 8. Okubo, A.: Diffusion and Ecological Problems: Mathematical Models (1980)

382 9. Turchin, P.: Quantitative Analysis of Movement: Measuring and Modeling Population Redistribution in Animals

383   and Plants. Sinauer Associates, Sunderland, MA, USA (1998)

384 10. Patterson, T.A., Basson, M., Bravington, M.V., Gunn, J.S.: Classifying movement behaviour in relation to

385   environmental conditions using hidden Markov models. Journal of Animal Ecology **78**(6), 1113–1123 (2009)

386 11. Schliehe-Diecks, S., Kappeler, P.M., Langrock, R.: On the application of mixed hidden Markov models to

387   multiple behavioural time series. Interface Focus **2**(2), 180–189 (2012). doi:10.1098/rsfs.2011.0077. Accessed

388   2014-05-02

389 12. Gurarie, E., Andrews, R.D., Laidre, K.L.: A novel method for identifying behavioural changes in animal

390   movement data. Ecology Letters **12**(5), 395–408 (2009)

391 13. Tracey, J.A., Zhu, J., Boydston, E., Lyren, L., Fisher, R.N., Crooks, K.R.: Mapping behavioral landscapes for

392   animal movement: a finite mixture modeling approach. Ecological Applications **23**(3), 654–669 (2012).

393   doi:10.1890/12-0687.1. Accessed 2015-04-20

394 14. van de Kerk, M., Onorato, D.P., Criffield, M.A., Bolker, B.M., Augustine, B.C., McKinley, S.A., Oli, M.K.:

395   Hidden semi-Markov models reveal multiphasic movement of the endangered Florida panther. Journal of

396   Animal Ecology **84**(2), 576–585 (2015)

397 15. McKellar, A.E., Langrock, R., Walters, J.R., Kesler, D.C.: Using mixed hidden Markov models to examine

398   behavioral states in a cooperatively breeding bird. Behavioral Ecology, 171 (2014). doi:10.1093/beheco/aru171.

399   Accessed 2015-04-21

400 16. Dean, B., Freeman, R., Kirk, H., Leonard, K., Phillips, R.A., Perrins, C.M., Guilford, T.: Behavioural mapping

401   of a pelagic seabird: combining multiple sensors and a hidden Markov model reveals the distribution of at-sea

402   behaviour. Journal of the Royal Society Interface, 20120570 (2012)

403   17. Morales, J.M., Haydon, D.T., Frair, J., Holsinger, K.E., Fryxell, J.M.: Extracting more out of relocation data:

404   building movement models as mixtures of random walks. Ecology **85**(9), 2436–2445 (2004)

405   18. Franke, A., Caelli, T., Kuzyk, G., Hudson, R.J.: Prediction of wolf (*Canis lupus*) kill-sites using hidden Markov

406   models. Ecological Modelling **197**(1–2), 237–246 (2006). doi:10.1016/j.ecolmodel.2006.02.043. Accessed

407   2015-04-29

408   19. Dean, B., Freeman, R., Kirk, H., Leonard, K., Phillips, R.A., Perrins, C.M., Guilford, T.: Behavioural mapping

409   of a pelagic seabird: combining multiple sensors and a hidden Markov model reveals the distribution of at-sea

410   behaviour. Journal of The Royal Society Interface **10**(78), 20120570 (2013). doi:10.1098/rsif.2012.0570.

411   Accessed 2016-06-07

412   20. Richards, S.A.: Testing ecological theory using the information-theoretic approach: examples and cautionary

413   results. Ecology **86**(10), 2805–2814 (2005)

414   21. Burnham, K.P., Anderson, D.R.: Model Selection and Inference: A Practical Information-Theoretic Approach.

415   Springer, New York (1998)

416   22. Zucchini, W., MacDonald, I.L.: Hidden Markov Models for Time Series: An Introduction Using R. CRC Press,

417   ??? (2009)

418   23. Visser, I., Speekenbrink, M.: depmixS4: An R package for hidden Markov models. Journal of Statistical

419   Software **36**(7), 1–21 (2010)

420   24. R Core Team: R: A Language and Environment for Statistical Computing. R Foundation for Statistical

421   Computing, Vienna, Austria (2015). R Foundation for Statistical Computing. https://www.R-project.org/

422   25. Biernacki, C., Celeux, G., Govaert, G.: Assessing a mixture model for clustering with the integrated completed

423   likelihood. IEEE transactions on pattern analysis and machine intelligence **22**(7), 719–725 (2000)

424   26. Potts, J.R., Auger-Méthé, M., Mokross, K., Lewis, M.A.: A generalized residual technique for analysing

425   complex movement models using earth mover's distance. Methods in Ecology and Evolution **5**(10), 1012–1022

426   (2014). Accessed 2016-06-07

427   **Figures**

Figure 1 Simulation Results BIC-optimal state frequency for 2-6 state HMMs with and without covariate transition on 100 two-state hidden Markov models with covariate transitioning simulations.

Figure 2 Sample figure title. Figure legend text.