

## RESEARCH

# Incorporating Periodic Variability in Hidden Markov Models for Animal Movement

Michael Li<sup>1\*</sup> and Benjamin M. Bolker<sup>1,2</sup>

\*Correspondence:

lim88@mcmaster.ca

<sup>1</sup> Department of Biology,

McMaster University, 1280 Main

St. West, L8S 4K1, Hamilton,

Ontario, Canada

Full list of author information is  
available at the end of the article

## 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 probabilities of group occupancy, or 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 long-term 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 analyses of panther data showed that model misspecification (omitting important sources of variation) can lead to overfitting/overestimating the underlying number of behavioural states. Models incorporating temporal heterogeneity identify fewer underlying states, and can make out-of-sample predictions that capture observed diurnal and autocorrelated movement patterns exhibited by Florida panthers.

## Conclusion:

Incorporating temporal heterogeneity improved goodness of fit and predictive capability as well as reducing the selected number of behavioural states to a more biologically interpretable level. Our results suggest that incorporating additional structure in statistical models of movement behaviour 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 parsimonious description that can be used to understand past movements and predict future movements. Ecologists have long considered the effects of individual-level covariates (sex, age, nutritional status) and environmental covariates (habitat type, location of predators or prey) on movement [1–3]. More recently, modelers have developed *hidden Markov models* (HMMs) [4–6] — used in animal ecology under the rubric of

the “multiphasic movement framework” [7] — that consider the effects of organisms’ *internal* states; in particular, HMMs model animal movement as though individual animals’ movement behaviour at particular times is determined by which of a discrete set of unobserved movement states (e.g. “foraging”, “traveling”, “resting”) they currently occupy. Conditional on the state occupied by an individual, HMMs typically assume that animals follow a correlated random walk model [8, 9].

Ever-increasing capabilities of remote sensors are making movement data available over an ever-wider range of time scales, at both higher resolution (e.g. hourly data from GPS collars vs. daily or weekly fixes for radio or VHF collars) and longer extent (e.g. from a few days to months or years). When analyzing such long-term data, ecologists will more often have to account for temporal variability in movement behaviour at diurnal and seasonal scales that were previously not captured in the data.

HMMs have typically been used to model movements over short time scales, where the probability of transitioning between movement states is approximately constant. Changes in transition probabilities based on the local environment can be accounted for by incorporating environmental covariates in the HMM [10], or inferred from direct comparisons between inferred states and environmental conditions [7]. Schliehe-Diecks et al. [11] considered temporal trends in behavioural transitions over the time scales of a six-hour observation period, but in general ecologists have turned to other tools to describe behavioural changes over longer (diurnal, seasonal, or ontogenetic) time scales [12].

For movement behaviours that change on a fast time scale, such that movement behaviours recorded at successive observations are effectively independent, *finite mixture models* (FMMs) — which can be considered a special case of HMMs where the probability of state occupancy is independent of the previous state — can adequately describe movement [13]. When movement varies over long time scales

(relative to the time between observations) with little short-term persistence or correlation, movement could be well represented by FMMs where the occupancy probabilities change deterministically over time. Thus FMMs and HMMs, with or without temporal variation in the occupancy or transition probabilities, form a useful family of models for capturing changes in movement behaviour over a range of time scales.

Our primary goal in this paper is to introduce the use of HMMs with temporally varying transition probabilities – in particular, transition probabilities that follow a diurnal cycle – for modeling animal movement recorded over long time scales. In addition to simulation-based examples, we also re-analyze data from van de Kerk et al. [14], who used temporally homogeneous hidden semi-Markov models (HSMMs: an extension of HMMs that allow flexible modelling of the distribution of *dwell times*, the lengths of consecutive occupancy of a behavioural state) to describe the movement and putative underlying behavioural states of Florida panthers (*Puma concolor coryi*).

van de Kerk et al. [14] found that the best-fitting HSMMs incorporated a surprisingly large number of hidden behavioural states (as many as six for individuals with a large amount of available data); for reasons of computational practicality and biological interpretability, they restricted their detailed analysis to models with only three underlying states. In contrast, most studies using HMM have chosen the number of underlying states *a priori*, typically using either two [6, 7, 11, 15], or three states [16–18]. In contrast, Dean et al. [16] evaluated models with up to 10 states, but like van de Kerk et al. they chose to consider only models with three states. As van de Kerk et al. [14] comment, and as we discuss further below, behavioural repertoires with more than three distinct states are difficult to interpret — one 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 allows us to select models with fewer latent states; we also fit models to simulated data with varying numbers of latent states, and with and without 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 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 considering multivariate (i.e., step length plus turning angle) observations. To check that this simplification does not distort our conclusions 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 multivariate, 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.)

van de Kerk et al. [14] used the Bayesian (Schwarz) information criterion (BIC) to test the relative penalized goodness of fit for models ranging from 2 to 6 latent

states. In general, BIC values decreased as the number of states increased from three to six states, suggesting that the six-state model was favoured statistically; however, the authors used three-state models in most of their analyses for ease of biological interpretation. We follow van de Kerk et al. [14] in using BIC-optimality (i.e., minimum BIC across a family of models) as the criterion for identifying the best model, because we are interested in explaining the data generation process by identifying the “true” number of underlying movement states.

Using BIC also simplifies evaluation of model selection procedures; it is easier to test whether our model selection procedure has selected the model used to simulate the data, rather than testing whether it has selected the model with the minimal Kullback-Leibler distance [19]. We recognize that ecologists will often be interested in maximizing predictive accuracy rather than selecting a true model, and that as usual in ecological systems the true model will be far more complicated than any candidate model [20]. We have repeated some of our analyses using AIC rather than BIC (not shown); for our examples, the qualitative conclusions stated here for BIC-optimality carry over to analyses using AIC.

#### Model description

In a HMM, the joint likelihood of *emissions* (i.e., direct observations)  $\mathbf{Y} = \mathbf{y}_1, \dots, \mathbf{y}_T$  and a hidden state sequence  $\mathbf{Z}, z_t \in \{1, \dots, n\}, t = 1, \dots, T$ , given model parameters  $\boldsymbol{\theta}$  and covariates  $\mathbf{X}_{1:T} = \mathbf{x}_1, \dots, \mathbf{x}_T$ , can be written as:

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

138 The emissions  $\mathbf{y}_i$  are boldfaced to denote that we may have a vector of observations  
 139 at each time point (e.g., step length and turning angle). The model contains three  
 140 distinct components:

141 **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  
 142  $t = 1$  given that the covariates are  $\mathbf{x}_1$ , times the vector of observations  $\mathbf{y}_1$   
 143 conditioned on state  $z_1$  and covariates  $\mathbf{x}_1$ .

144 **Transition probability**  $P(z_k = j | z_{k-1} = i, \mathbf{x}_k)$ : the probability of a transition  
 145 from state  $i$  at time  $t = k - 1$  to state  $j$  at time  $t = k$ , given covariates  $\mathbf{x}_k$ .

146 **Emission probability**  $P(\mathbf{y}_k | z_k, \mathbf{x}_k)$ : a vector of observations  $\mathbf{y}_k$  given state  $z_k$  at  
 147 time  $t = k$  and covariates  $\mathbf{x}_k$ .

148 Eq. 1 gives the likelihood of the observed sequence given (conditional on) a partic-  
 149 ular hidden sequence. In order to calculate the overall, unconditional (or marginal)  
 150 likelihood of the observed sequence, we need to average over all possible hidden  
 151 sequences. There are several efficient algorithms for computing the marginal like-  
 152 lihood and numerically estimating parameters [21]; we used those implemented in  
 153 the `depmixS4` package for R [22, 23].

154 For an  $n$ -state HMM, we need to define an  $n \times n$  matrix that specifies the proba-  
 155 bilities  $\pi_{ij}$  of being in movement states  $j$  at time  $t + 1$  given that the individual is  
 156 in state  $i$ . The FMM is a special case of HMM where the probabilities of *entering*  
 157 a given state are identical across all states — i.e., the probability of occupying a  
 158 state at the next time step is independent of the current state occupancy. It can be  
 159 modelled in the HMM framework by setting the transition probabilities  $\pi_{ij} = \pi_{i*}$ .

160 In any case, the transition matrix  $\pi_{ij}$  must respect the constraints that (1) all  
 161 probabilities are between 0 and 1 and (2) transition probabilities out of a given state  
 162 sum to 1. As is standard for HMMs with covariates [22], we define this multinomial  
 163 logistic model in terms of a linear predictor  $\eta_{ij}$ , where  $\eta_{i1}$  is set to 1 without loss

of generality (i.e. we have only  $n \times (n - 1)$  distinct parameters; we index  $j$  from 2 to  $n$  for notational clarity):

$$\begin{aligned}\pi_{ij} &= \exp(\eta_{ij}(t)) / \left( 1 + \sum_{j=2}^n \exp(\eta_{ij}(t)) \right), \text{ for } j = 2, \dots, n \\ \pi_{i1} &= 1 - \sum_{j=2}^n \pi_{ij}\end{aligned}\tag{2}$$

We considered four different transition models for diurnal variation in behaviour, incorporating hour-of-day as a covariate following the general approach of Morales et al. [17] of incorporating covariate dependence in the transition matrix.

**Multiple block transition** Here we assume piecewise-constant transition probabilities. The transition probability  $\pi_{ij}$  is a function of time (hour of day), where it is assigned to one of  $M$  different time blocks:

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

where  $a_{ijm}$  are parameters, and  $\delta_{m=t}$  is a Kronecker delta ( $\delta_{m=t} = 1$  for the time block corresponding to time  $t$ , and 0 otherwise).

**Quadratic transition model** We assume the elements of the linear predictor are quadratic functions of hour:

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

The quadratic model is not diurnally continuous, i.e. there is no constraint that forces  $\eta_{ij}(0) = \eta_{ij}(24)$ ; imposing a diurnal continuity constraint would collapse the model to a constant.



179 **Sinusoidal transition model** A sinusoidal model with a period of 24 hours is  
 180 identical in complexity to the quadratic model, but automatically satisfies  
 181 the diurnal continuity constraint:

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

182 **Hourly model** Lastly, we extended the multi-block approach and assign a differ-  
 183 ent transition matrix for every hour of the day. This model is included for  
 184 comparative purposes; due to the large number of parameters in the model  
 185 (more than  $24n(n-1)$  for a HMM with  $n$  states), it is not really practical.  
 186 We only fitted up to four states using the hourly model.

187 Other periodic functions, such as Fourier series (i.e., the sinusoidal transition  
 188 model augmented by additional sinusoidal components at higher frequencies) or  
 189 periodic splines, could also be considered.

190 Model complexity and the number of parameters increase as the number of la-  
 191 tent states increase. For a fixed number of states homogeneous FMMs are simplest,  
 192 followed by homogeneous HMMs and finally by FMMs and HMMs incorporating  
 193 temporal heterogeneity. In general, the number of free parameters in an HMM is the  
 194 sum of the number of free parameters for each of the three model components (initial  
 195 states, transition probabilities, and emissions). Let  $n$  be the number of hidden states  
 196 and  $k_i, k_t, k_e$  be the number of parameters describing the covariate-dependence of  
 197 the prior distribution, transition function and emission distributions; that is, for a  
 198 homogeneous model,  $k = 1$ , while a single numeric covariate or a categorical predic-  
 199 tor with two levels would give  $k = 2$ . Then the number of free parameters of an HMM  
 200 is: *[Initial states]*  $k_i \cdot (n-1) +$  *[Transition probabilities]*  $k_t \cdot n \cdot (n-1) +$  *[Emission*  
 201 *parameters]*  $k_e \cdot n$ . As the number of states increases, the number of free parameters  
 202 in (homogeneous or heterogeneous) FMMs and time-homogeneous HMMs increases

linearly, whereas for HMMs with temporal heterogeneity (or covariate-dependent transitions more generally) the number increases quadratically.

#### Model evaluation

We used the `depmixS4` package [22] to fit covariate-dependent transition HMMs, simulate states and step lengths using the estimated parameters, and estimate the most likely states with the Viterbi algorithm.

We ran a simulation experiment in which we fitted HMMs with both homogeneous and heterogeneous transition probabilities to simulated data with heterogeneous transition probabilities, to see whether the correct (heterogeneous-transition) model correctly identified the number of states while the misspecified (homogeneous-transition) model overestimated the number of states. We used 100 realizations of a two-state HMM with sinusoidally varying transition probabilities and fitted it with HMMs ranging from 2 to 5 hidden states, with and without temporal heterogeneity in the transition probabilities.

We used three approaches to assess the fit of both time-homogeneous and time-inhomogeneous HMMs with 3 to 6 states to step-length data from the four of the thirteen Florida panthers with the most data ( $> 9000$  observations). (1) BIC was used to compare the goodness of fit of each model type. The model with the lowest BIC was selected to be the optimal-BIC model and all BICs were adjusted to  $\Delta\text{BIC}$  based on the optimal-BIC model ( $\Delta\text{BIC} = \text{BIC} - \min(\text{BIC})$ ). (2) Comparing average step-length by hour of day for the observed data and for data simulated from the models shows how well a particular class of models can capture diurnal variation in behaviour. (3) Comparing temporal autocorrelations for the observed data and for data simulated from the models shows how well a particular class of models can capture serial correlation at both short and long time scales. As a complement, we also fitted FMM and FMM with priors on state occupancy that varied sinusoidally over time to compare the temporal effects in goodness of fit. 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.

We used simulations to predict expected hourly step lengths and autocorrelation functions (ACF). While the computation of expected step length and ACF is straightforward for FMMs, and feasible for homogeneous HMMs, the interaction between the geometric dwell time within each state and the temporally varying interaction probabilities makes it infeasible for more complex models. We used this approach to validate our models, comparing our simulated predictions with the observed movements. The more usual approach, generating predictions from the expected step lengths conditional on the most likely state sequence predicted by the Viterbi algorithm or pseudo-residuals [6, 21], is problematic because the predictions by these methods rely on the observed data. This approach 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 HMM models with different degrees of structural complexity.

## Results

The simulation experiment supports our hypothesis that homogeneous transition HMMs can overestimate the number of hidden states (Figure 1). Heterogeneous transition models can always predict the correct number of states (in 100/100 simulations, BIC correctly identifies  $n = 2$  as the number of states) whereas the temporally homogeneous models overestimate the number of states (the correct value,  $n = 2$ , is chosen most often, but in fewer than half of the simulations; values up to  $n = 5$  are frequently chosen).

[BMB: caption needs a little work]

256 The BIC-optimal number of states for time-homogeneous models is consistent  
 257 with van de Kerk et al.'s [14] results. For time-homogeneous models, the Weibull-  
 258 wrapped-Cauchy [14], Weibull-von Mises, and log Normal without turning angles all  
 259 identify the same BIC-optimal number of states. While the number of states identi-  
 260 fied by homogeneous-HMM models varies according to the emissions distributions  
 261 chosen, ranging from  $n = 5$  for Weibull steps alone to  $n = 7$  for the log Normal-  
 262 von Mises emissions model, the number of states identified by heterogeneous-HMM  
 263 models is consistent among emissions models ( $n = 5$ : Figure 2).

264 Models with temporal heterogeneity provide better fits to the data (lower  
 265 BIC) than homogeneous models in both FMM and HMM frameworks, but time-  
 266 homogeneous HMMs are better than FMMS with sinusoidal temporal heterogeneity  
 267 (Figure 3). Turning to the temporally heterogeneous HMMs (Figure 3, right panel),  
 268 we see that the model with different transition probabilities for each hour of the day  
 269 (HMM + THhourly) is overparameterized; it underperforms homogeneous HMM  
 270 with even 3 states, and gets much worse with 4 states. The multiple-block model  
 271 gives approximately the same BIC as the homogeneous HMM, although it gives the  
 272 BIC-optimal number of states as 4, in contrast to 6 for the homogeneous HMM.  
 273 Finally, the quadratic and sinusoidal models are the best models tested by far; they  
 274 both give the BIC-optimal number of states as 5, and they have similar goodness  
 275 of fit. However, the similarity between the quadratic and sinusoidal models may  
 276 be overstated in Figure 3 due to the very large variation in BIC (over thousands  
 277 of units) across the full range of models; the best-fit sinusoidal ( $n = 5$ ) model is  
 278 approximately 80 BIC units better than the best quadratic model (also  $n = 5$ ),  
 279 which would normally be interpreted as an enormous improvement in goodness of  
 280 fit (both models have 90 parameters).

281 The average hourly step lengths from the observed panther data exhibit a clear  
 282 diurnal pattern (Figure 4). As expected, temporally homogeneous models (whether

283 FMM or HMM) predict the same mean step length regardless of time of day, failing  
 284 to capture the diurnal activity cycle. All of the models incorporating temporal het-  
 285 erogeneity, including the temporally heterogeneous FMM, can capture the observed  
 286 patterns. However, the block model does markedly worse than the other tempo-  
 287 ral models (changing the block definitions might help by re-clustering/grouping  
 288 different hours or increasing the number of blocks), and the (overparameterized)  
 289 hourly model does better than any other model at capturing the early-evening peak  
 290 (but worse at capturing the mid-day trough). We also included average hourly step  
 291 lengths from three-state temporally homogeneous HMM Viterbi prediction to illus-  
 292 trate within sample predictions can capture the diurnal patterns, but fail to capture  
 293 out of sample predictions.

294 Like the diurnal pattern (Figure 4), the strong autocorrelation of the observed  
 295 step lengths at a 24-hour lag (Figure 5) shows the need to incorporate temporal  
 296 heterogeneity in the model — we could have reached this conclusion even without  
 297 developing any of the temporal-heterogeneity machinery. In contrast to the hourly  
 298 averages, the autocorrelation (ACF) captures both short- and long-term temporal  
 299 effects. HMM without temporal heterogeneity captures the short-term autocorre-  
 300 lation, but misses the long-term autocorrelation beyond a 7-hour lag. Temporally  
 301 homogeneous FMMs, by definition, produce no autocorrelation (neither short- nor  
 302 long-term autocorrelation). FMMs without temporal heterogeneity, although they  
 303 capture the diurnal pattern well, underpredict the degree of short-term autocorre-  
 304 lation.

305 The estimated emission parameter values (mean and standard deviation of the  
 306 step length in each state) are similar, for both homogeneous and heterogeneous  
 307 models, across all cats (Figure 7 shows a subset of cats). In general, the states  
 308 with longer mean step lengths are similar between homogeneous and heterogeneous  
 309 models. For cats 14 and 15, the states with the longest or next-longest mean step

lengths have similar means and standard deviations; for cats 1 and 2, three long-step states in the homogeneous HMM appear to divide two long-step states in the heterogeneous HMM. For short-step states, the heterogeneous HMM tends to identify a high-variance state, while the homogeneous HMM picks up states with very short step lengths (questionable in any case because we have not taken any special efforts to account 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, i.e. they will “discover” large number of states that are difficult to interpret biologically.

As usual, the appropriate approach depends on the goal of the analysis. If ecologists simply want to identify states and associate them with environmental characteristics, it might be sufficient to use a simple (homogeneous) HMM model, pre-specifying the number of states to a biologically sensible value, and then match *post hoc* Viterbi estimates of state occupancy with environmental variation in space and time [7].

On the other hand, if the goal of analysis is to make out-of-sample predictions about animal behaviour, such as in a management context, it is necessary to fit a covariate-dependent model that explicitly incorporates the switching process. While the Viterbi algorithm can be applied to work backward from observed behaviour to variations in state occupancy with environmental conditions even when using a homogeneous model, a homogeneous model can never *predict* behaviour that varies with environmental conditions.

337 Finally, if our goal is actually to estimate the number of discrete behavioural  
 338 modes underlying observed behaviour — keeping in mind that these discrete be-  
 339 havioural states are certainly an oversimplified representation of animals’ real in-  
 340 ternal states — then, as we have shown above, covariate-dependent models will  
 341 generally be required to avoid overestimating the number of states. We note that  
 342 researchers in cluster analysis (of which HMMs are a special case) have shown that  
 343 the technical conditions required for BIC to apply may be violated [24]; however,  
 344 BIC can be useful as an approximate *upper* limit on the number of states. Various  
 345 solutions to this problem have been proposed, including the “integrated classifica-  
 346 tion likelihood” [24], as well as a simpler “knee point” method [25] that looks for  
 347 the cluster size that corresponds to the largest change in BIC rather than to the  
 348 smallest overall BIC. [16] took a similar approach, but based on the log-likelihood  
 349 curve rather than the BIC. However, at least in our simulations the BIC does cor-  
 350 rectly identify the number of states when appropriate heterogeneity is included in  
 351 the model; the best practical criterion for identifying numbers of states underlying  
 352 animal movement data is an open question.

353 **[BMB: mike, can we take a quick try at computing ICL for our**  
 354 **data/sims?]**

$$\begin{aligned}
 ICL(m, K) = & \underbrace{\max_a \log f(x|\bar{z}, m, K, a)}_{\text{max log-likelihood}} - \frac{\lambda_{m,K}}{2} \log n \\
 & + \log \Gamma(K/2) + \sum_{k=1}^K \log \Gamma(\bar{n}_k + 1/2) \\
 & - K \log \Gamma(1/2) - \log \Gamma(n + K/2)
 \end{aligned}$$

355 ( $K$ =number of clusters;  $n$ =number of obs.;  $\lambda_{m,K}$  is “the number of free com-  
 356 ponents in  $a$  (i.e., number of free parameters: p. 8);  $\bar{n}_k$  is, I think, the num-  
 357 ber of data points which have their max. probability of being in cluster  $k$  [i.e.,

we should be able to get this at least approximately from Viterbi, although I'm not sure whether it's more appropriate to use Viterbi or an "instantaneous" max probability]). The `rebmix` package has an ICL criterion for its particular objects (`sos::findFn("{integrated classification likelihood}")`), but I don't know if it'll be useful. I think Raftery and Fraley also have a modified IC for clusters?

We have shown the value of incorporating temporal heterogeneity in animal movement models, but much remains to be done. Our model neglects other predictors — such as sex, habitat type or location with respect to environmental features such as roads — that can potentially improve predictive accuracy and further reduce the estimated number of states. While adding more covariates is in principle straightforward, including even a reasonable range of biological complexity in a HMM with state-dependent transitions rapidly becomes intractable in both computational time and the complexity of model selection. 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. 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.

## 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.

#### Acknowledgements

We would like to thank Madelon van de Kerk, Madan Oli, and David Onorato for their previous work on Florida panthers. We also would like to thank McMaster University, Florida Fish and Wildlife Conservation Commission and many individuals for data collection and fieldwork. Lastly, we thank Madelon van de Kerk for making the data available at the Institutional Repository at the University of Florida (IR@UF).

#### Ethics approval

All data used are secondary, drawn from an existing institutional data repository.

#### Consent for publication

Not applicable.

#### Funding

This study was funded by NSERC Discovery Grant 386590-2010 to BMB.

#### 1 Data accessibility

Hourly step lengths and turning angles of male and female Florida panthers available at <http://ufdc.ufl.edu//IR00004241/00001>.

#### Authors' contributions

ML designed analyses and simulations; ran analyses and simulations; and co-wrote the text of the paper. BMB designed analyses and simulations and co-wrote the text of the paper.

#### Competing interests

The authors declare that they have no competing interests.

#### Author details

<sup>1</sup> Department of Biology, McMaster University, 1280 Main St. West, L8S 4K1, Hamilton, Ontario, Canada. <sup>2</sup> Department of Mathematics and Statistics, McMaster University, 1280 Main St. West, L8S 4K1, Hamilton, Ontario, Canada.

#### References

- Patterson, T.A., Thomas, L., Wilcox, C., Ovasikainen, O., Matthiopoulos, J.: State-space models of individual animal movement. *Trends in Ecology & Evolution* **23**(2), 87–94 (2008)
- McKenzie, H.W., Lewis, M.A., Merrill, E.H.: First passage time analysis of animal movement and insights into the functional response. *Bulletin of Mathematical Biology* **71**(1), 107–129 (2009)
- Pal, S., Ghosh, B., Roy, S.: Dispersal behaviour of free-ranging dogs (*Canis familiaris*) in relation to age, sex, season and dispersal distance. *Applied Animal Behaviour Science* **61**(2), 123–132 (1998)
- Firle, S., Bommarco, R., Ekbohm, B., Natiello, M.: The influence of movement and resting behavior on the range of three carabid beetles. *Ecology* **79**(6), 2113–2122 (1998). doi:10.1890/0012-9658(1998)079[2113:TIOMAR]2.0.CO;2. Accessed 2015-04-14

- 422 5. Nathan, R., Getz, W.M., Revilla, E., Holyoak, M., Kadmon, R., Saltz, D., Smouse, P.E.: A movement ecology  
423 paradigm for unifying organismal movement research. *Proceedings of the National Academy of Sciences*  
424 **105**(49), 19052–19059 (2008). doi:10.1073/pnas.0800375105. Accessed 2015-04-29
- 425 6. Langrock, R., King, R., Matthiopoulos, J., Thomas, L., Fortin, D., Morales, J.M.: Flexible and practical  
426 modeling of animal telemetry data: hidden Markov models and extensions. *Ecology* **93**(11), 2336–2342 (2012).  
427 doi:10.1890/11-2241.1. Accessed 2013-10-24
- 428 7. Fryxell, J.M., Hazell, M., Börger, L., Dalziel, B.D., Haydon, D.T., Morales, J.M., McIntosh, T., Rosatte, R.C.:  
429 Multiple movement modes by large herbivores at multiple spatiotemporal scales. *Proceedings of the National*  
430 *Academy of Sciences* **105**(49), 19114–19119 (2008). doi:10.1073/pnas.0801737105. Accessed 2013-04-09
- 431 8. Okubo, A.: *Diffusion and Ecological Problems: Mathematical Models* (1980)
- 432 9. Turchin, P.: *Quantitative Analysis of Movement: Measuring and Modeling Population Redistribution in Animals*  
433 *and Plants*. Sinauer Associates, Sunderland, MA, USA (1998)
- 434 10. Patterson, T.A., Basson, M., Bravington, M.V., Gunn, J.S.: Classifying movement behaviour in relation to  
435 environmental conditions using hidden Markov models. *Journal of Animal Ecology* **78**(6), 1113–1123 (2009)
- 436 11. Schliehe-Diecks, S., Kappeler, P.M., Langrock, R.: On the application of mixed hidden Markov models to  
437 multiple behavioural time series. *Interface Focus* **2**(2), 180–189 (2012). doi:10.1098/rsfs.2011.0077. Accessed  
438 2014-05-02
- 439 12. Gurarie, E., Andrews, R.D., Laidre, K.L.: A novel method for identifying behavioural changes in animal  
440 movement data. *Ecology Letters* **12**(5), 395–408 (2009)
- 441 13. Tracey, J.A., Zhu, J., Boydston, E., Lyren, L., Fisher, R.N., Crooks, K.R.: Mapping behavioral landscapes for  
442 animal movement: a finite mixture modeling approach. *Ecological Applications* **23**(3), 654–669 (2012).  
443 doi:10.1890/12-0687.1. Accessed 2015-04-20
- 444 14. van de Kerk, M., Onorato, D.P., Criffield, M.A., Bolker, B.M., Augustine, B.C., McKinley, S.A., Oli, M.K.:  
445 Hidden semi-Markov models reveal multiphasic movement of the endangered Florida panther. *Journal of*  
446 *Animal Ecology* **84**(2), 576–585 (2015)
- 447 15. McKellar, A.E., Langrock, R., Walters, J.R., Kesler, D.C.: Using mixed hidden Markov models to examine  
448 behavioral states in a cooperatively breeding bird. *Behavioral Ecology*, 171 (2014). doi:10.1093/beheco/aru171.  
449 Accessed 2015-04-21
- 450 16. Dean, B., Freeman, R., Kirk, H., Leonard, K., Phillips, R.A., Perrins, C.M., Guilford, T.: Behavioural mapping  
451 of a pelagic seabird: combining multiple sensors and a hidden Markov model reveals the distribution of at-sea  
452 behaviour. *Journal of the Royal Society Interface*, 20120570 (2012)
- 453 17. Morales, J.M., Haydon, D.T., Frair, J., Holsinger, K.E., Fryxell, J.M.: Extracting more out of relocation data:  
454 building movement models as mixtures of random walks. *Ecology* **85**(9), 2436–2445 (2004)
- 455 18. Franke, A., Caelli, T., Kuzyk, G., Hudson, R.J.: Prediction of wolf (*Canis lupus*) kill-sites using hidden Markov  
456 models. *Ecological Modelling* **197**(1–2), 237–246 (2006). doi:10.1016/j.ecolmodel.2006.02.043. Accessed  
457 2015-04-29
- 458 19. Richards, S.A.: Testing ecological theory using the information-theoretic approach: examples and cautionary  
459 results. *Ecology* **86**(10), 2805–2814 (2005)
- 460 20. Burnham, K.P., Anderson, D.R.: *Model Selection and Inference: A Practical Information-Theoretic Approach*.  
461 Springer, New York (1998)
- 462 21. Zucchini, W., MacDonald, I.L.: *Hidden Markov Models for Time Series: An Introduction Using R*. CRC Press,  
463 Boca Raton, FL, USA (2009)
- 464 22. Visser, I., Speekenbrink, M.: *depmixS4: An R package for hidden Markov models*. *Journal of Statistical*

465       Software **36**(7), 1–21 (2010)

466   23. R Core Team: R: A Language and Environment for Statistical Computing. R Foundation for Statistical

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

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

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

470   25. Zhao, Q., Xu, M., Fränti, P.: Knee Point Detection on Bayesian Information Criterion, pp. 431–438. IEEE, ???

471       (2008). doi:10.1109/ICTAI.2008.154.

472       <http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=4669805> Accessed 2016-06-21

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

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

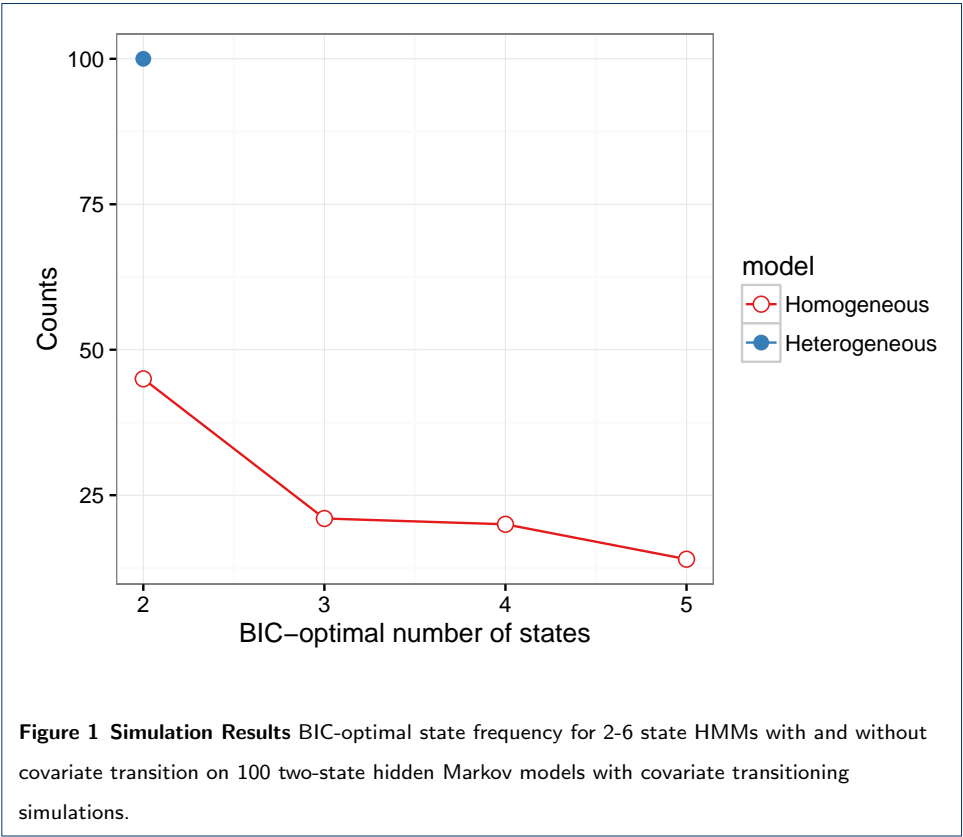
475       (2014). Accessed 2016-06-07

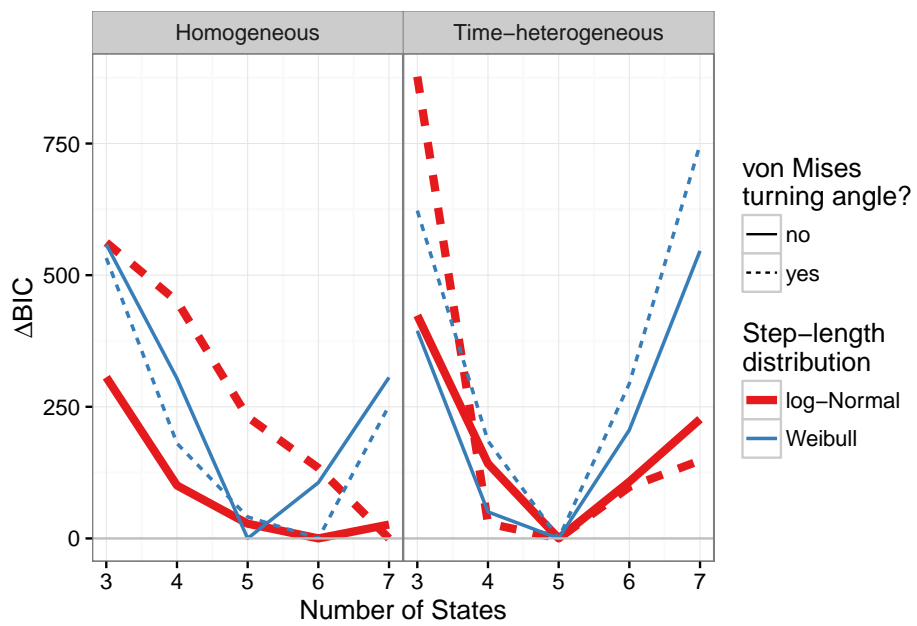
476   **Tables**

**Table 1** Cat ID and number of observations; ID numbers are given matching those shown by van de Kerk et al. 2015 and those in the data located at the UF Institutional repository (IR@UF).

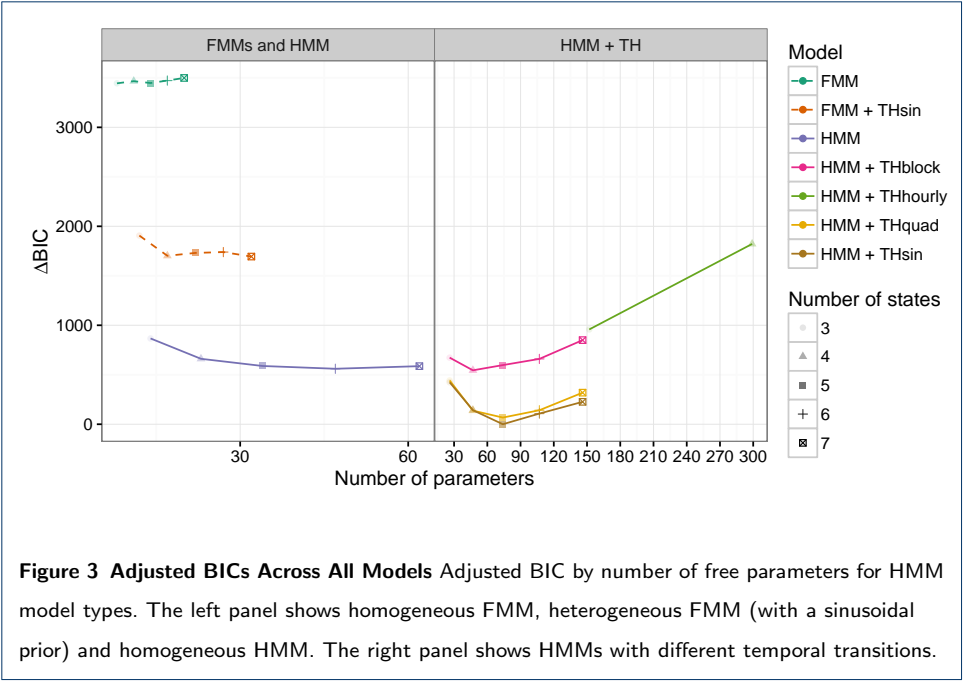
van de Kerk 2015	IR@UF	Number of Observations
130	1	10286
131	2	9458
48	14	14645
94	15	10250

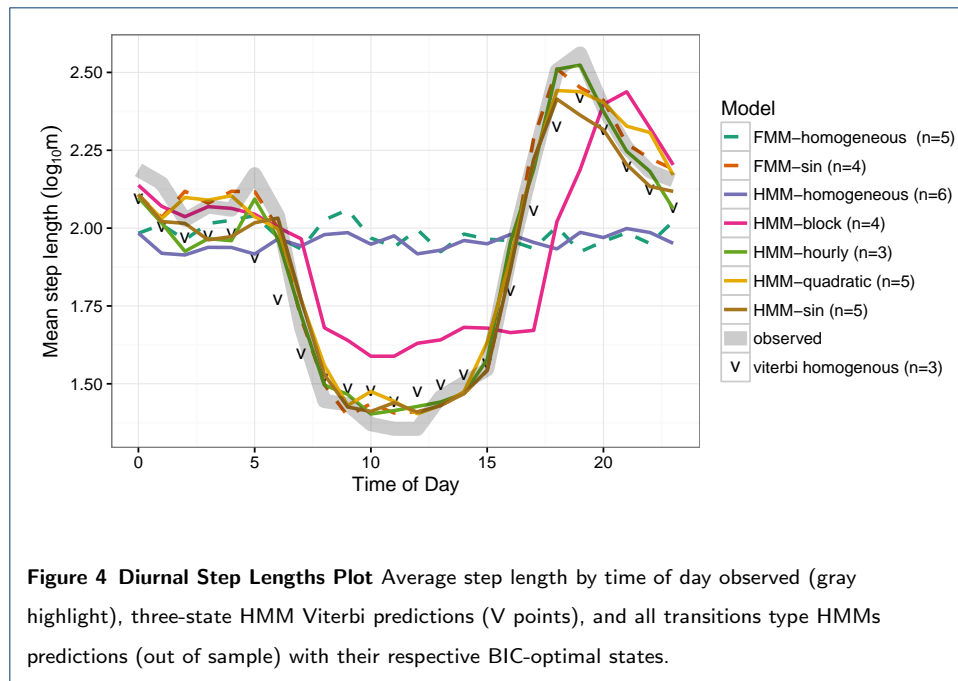
477   **Figures**

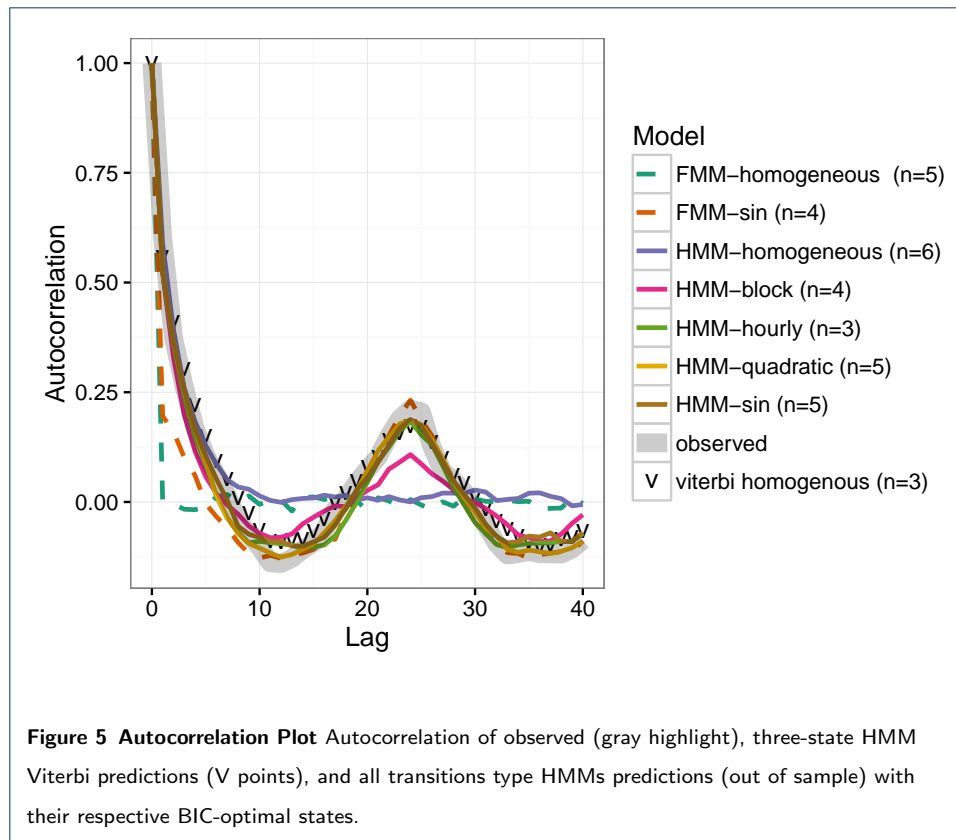


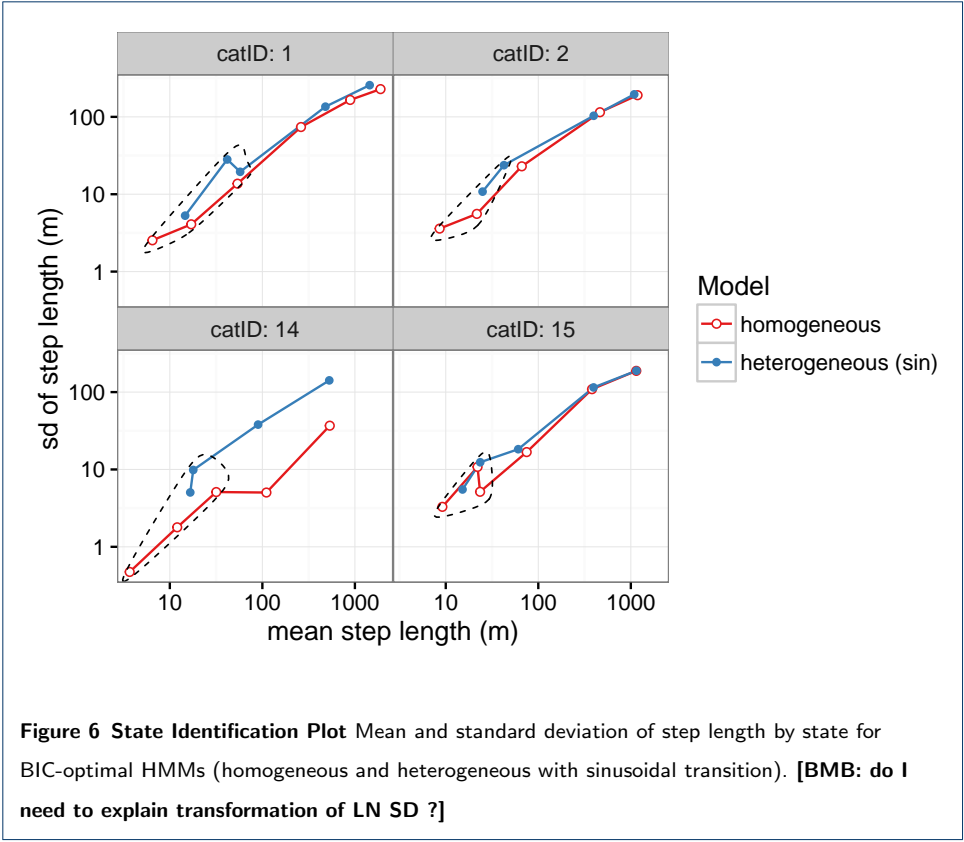


**Figure 2 Adjusted BIC Emission Distribution Comparison** Adjusted Bayesian information criterion values for 3-7 state HMMs with different step-length distributions, with and without temporal transitions and turning angles.









**Figure 6 State Identification Plot** Mean and standard deviation of step length by state for BIC-optimal HMMs (homogeneous and heterogeneous with sinusoidal transition). [BMB: do I need to explain transformation of LN SD ?]