

Incorporating periodic variability in hidden Markov models for animal movement

Michael Li and Benjamin M. Bolker

July 24, 2016

Abstract

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. More generally, we point out that model misspecification (omitting important sources of variation) can lead to overfitting, and as a corollary 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

1 Introduction

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 (McKenzie et al., 2009; Pal et al., 1998; Patterson et al., 2008). More recently, modelers have developed *hidden Markov models* (HMMs) (Firle et al., 1998; Langrock et al., 2012; Nathan et al., 2008) — used in animal ecology under the rubric of the “multiphasic movement framework” (Fryxell et al., 2008) — 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 standard correlated walk model (Okubo, 1980; Turchin, 1998).

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 significant fractions of a year, or longer). 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 latent/hidden behavioural state/mode transition probabilities based on the local environment can be accounted for incorporating environmental covariates in the HMM (Patterson et al., 2009), or by more di-

rect comparisons between inferred states and environmental conditions (Fryxell et al., 2008). Schliehe-Diecks et al. (2012) consider 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 (Gurarie et al., 2009).

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 (Tracey et al., 2012). 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. (2015), 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. (2015) 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 (Fryxell et al., 2008; Langrock et al., 2012; McKellar et al., 2014; Schliehe-Diecks et al., 2012), or three states (Dean et al., 2012; Franke et al., 2006; Morales et al., 2004). In contrast, Dean et al. (2013) evaluated models with up to 10 states, but like van de Kerk et al. (2015) they chose to consider only models with three states (their criterion for choosing ...). As van de Kerk et al. (2015) 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.

check Zhao et al.
(2008)!

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.

2 Methods

2.1 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. (2015). 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) (Turchin, 1998; van de Kerk et al., 2015).

van de Kerk et al. (2015) used hidden semi-Markov models (HSMM), an extension of HMM that permits explicit modelling of dwell times (Langrock et al., 2012), considering both Poisson and negative binomial distributions for dwell times. As shown by van de Kerk et al. (2015) (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. (2015) 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. 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. (2015), 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. (2015) 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. (2015) 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

it occurs to me that we haven't really included any information about simulation testing in this version! We should probably add a short section; if don't, then we should delete following sentence

model used to simulate the data, rather than testing whether it has selected the model with the minimal Kullback-Leibler distance (Richards, 2005). 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 (Burnham and Anderson, 1998); we believe that the qualitative conclusions stated here for BIC-optimality will carry over to analyses using AIC instead.

2.2 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:

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

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

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 $t = 1$ where the covariate is \mathbf{x}_1 , times the vector of observations \mathbf{y}_1 conditioned on covariates \mathbf{x}_1 and state z_1 .

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

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

Eq. 1 gives the likelihood of the observed sequence given (conditional on) a particular hidden sequence. In order to calculate the overall, unconditional (or marginal) likelihood of the observed sequence, we need to average over all possible hidden sequences. There are several efficient algorithms for computing the marginal likelihood and numerically estimating parameters (Zucchini and MacDonald, 2009); we used those implemented in the `depmixS4` package for R (R Core Team, 2015; Visser and Speekenbrink, 2010).

maybe give the formula for number of parameters (constrained) for each model type here?...ML: $\text{Sum}(1+T+E)$, $I=\text{Prior}=n(s-1)$ where n is the number of parameters for the prior complexity, $T = ns(s-1)$, $E=2s$

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

In any case, the transition matrix π_{ij} must respect the constraints that (1) all probabilities are between 0 and 1 and (2) transition probabilities out of a given state sum to 1. As is standard for HMMs with covariates (Visser and Speekenbrink, 2010), we define this multinomial logistic model in terms of a linear predictor η_{ij} , where η_{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):

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

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

We considered four different transition models for diurnal variation in be-

haviour, incorporating hour-of-day as a covariate following the general approach of Morales et al. (2004) of incorporating covariate dependence in the transition matrix.

Multiple block transition Here we assume piecewise-constant transition probabilities. The transition probability π_{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 at the corresponding time t , and 0 otherwise).

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

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

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

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

Hourly model Lastly, we extended the multi-block approach and assign a different transition matrix for every hour of the day. This model is included for comparative purposes due to the large number of parameters in the

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

Other periodic functions, such as Fourier series (the sinusoidal transition model augmented by additional sinusoidal components at higher frequencies) or periodic splines, could be useful directions for future exploration.

2.3 Model evaluation

We used the `depmixS4` package 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 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) Comparing BICs to the optimal-BIC model within each type of transition complexity ($\Delta\text{BIC} = \text{BIC} - \min(\text{BIC})$) assesses the overall goodness of fit of each model type. (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 the 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 captures serial correlation at both short and long scales.

Model complexity and the number of parameters increase as the number of latent states increase, FMM to HMM, and lastly, FMM and HMM incorporating temporal heterogeneity. The number of free parameters in an HMM can be generalized by summing up the number of free parameters of the three distinct components. Let n be the number of hidden states and k_i, k_t, k_e be the number of parameters describing the covariate-dependence of the prior distribu-

tion, transition function and emission distributions; that is, for a homogeneous model, $k = 1$, while a single numeric covariate or a categorical predictor with two levels would give $k = 2$. Then the number of free parameters of an HMM is:

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

As the number of states increases, the number of free parameters in time-homogeneous FMMs and HMMs and FMMs with temporal heterogeneity will increase linearly, whereas HMMs with temporal heterogeneity will increase quadratically (Eq. 3). When comparing BICs, it is important to account for the tradeoff between log-likelihood and number of states, but also log-likelihood and number of free parameters.

We used simulations to predict hourly step length and ACF because, while the computation is reasonably straightforward for FMMs, and manageable for homogeneous HMMs, the interaction between the geometric dwell time within each state and the temporally varying interaction probabilities makes it unreasonably complex. We used this approach to validate our models and comparing these models with the 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 (Langrock et al., 2012; Zucchini and MacDonald, 2009), 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.

think about/check
out **pseudo-
residuals**

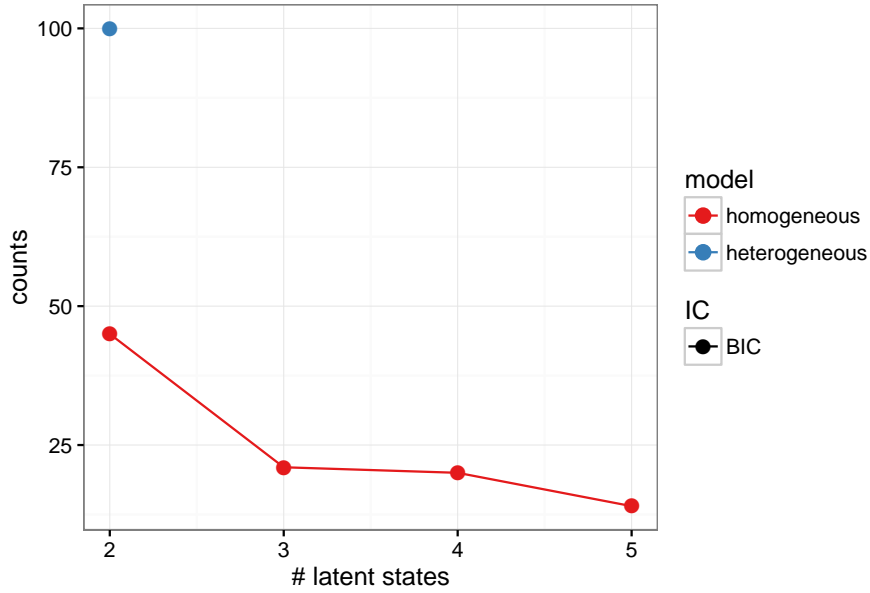


Figure 1: Simulation test for BIC-optimal approach

3 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. Figure 1 shows heterogeneous transition models can always predict the correct number of states, whereas, can overestimate the number of states via BIC-optimal approach.

The description of the simulations should be in Methods, not Results; here we should describe the results of the simulations only

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

is it **exactly** consistent? Weibull/vM/homog gets 5, not 6? Can we explain?

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

I took out "hourly HMM" here - more coherent to describe it above

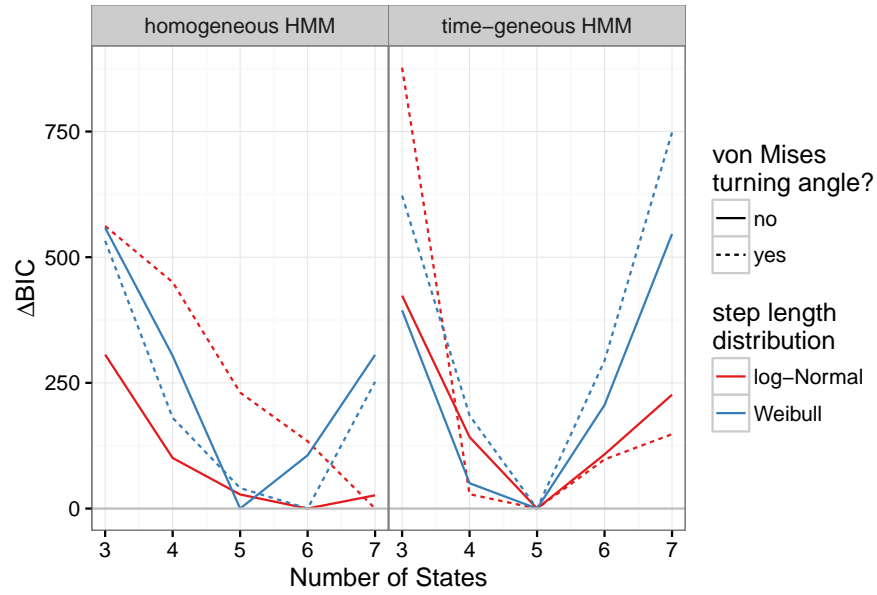


Figure 2: Change in BIC across models, cat #1. ΔBIC is scaled individually per model, so that $\Delta\text{BIC} = 0$ represents the BIC-optimal number of states for a particular model.

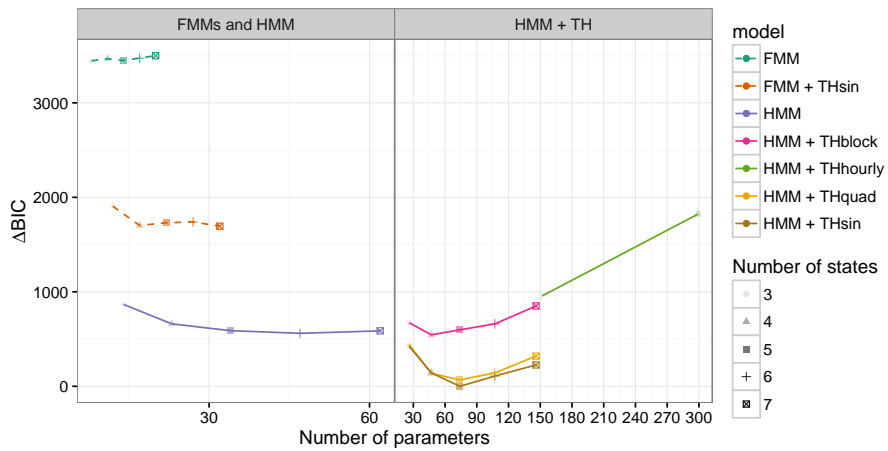


Figure 3: Overall adjusted BIC comparisons among models, cat #1. **todo:** use dashed lines for FMMs (leadup to next figure)?

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.

Figure 3 shows that models with temporal heterogeneity are better (lower BIC) than homogeneous models in both FMM and HMM frameworks, but time-homogeneous HMMs are better than FMMs with sinusoidal temporal heterogeneity. 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).

Figure 4 shows a clear diurnal pattern from the average hourly step lengths from the observed data. 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 (overparameterized) hourly model does better than any other model at captur-

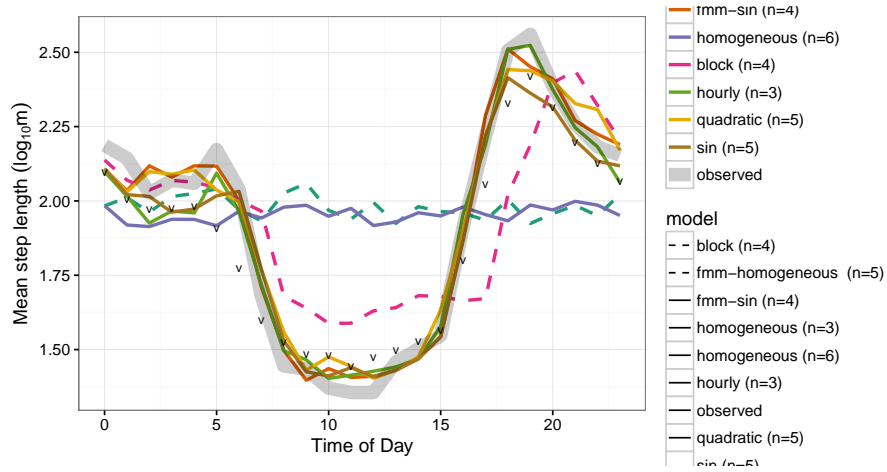


Figure 4: Average step length by time of day. **todo:** leave out hmm quad for clarity? fix factor labels; make y axis log-scale rather than showing log-step-length; make colours consistent with prev fig (may need scalecolourmanual [need underscores] with RColorBrewer::brewer.pal(10,'Dark2') or some such. Handle obs data as in Guelph talk. Make FMM lines dashed use manual linetype scale))

ing 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 shown in 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 hourly averages, the autocorrelation (ACF) captures both short- and long-term temporal effects. HMM without temporal heterogeneity captures the short-term autocorrelation, but misses the long-term

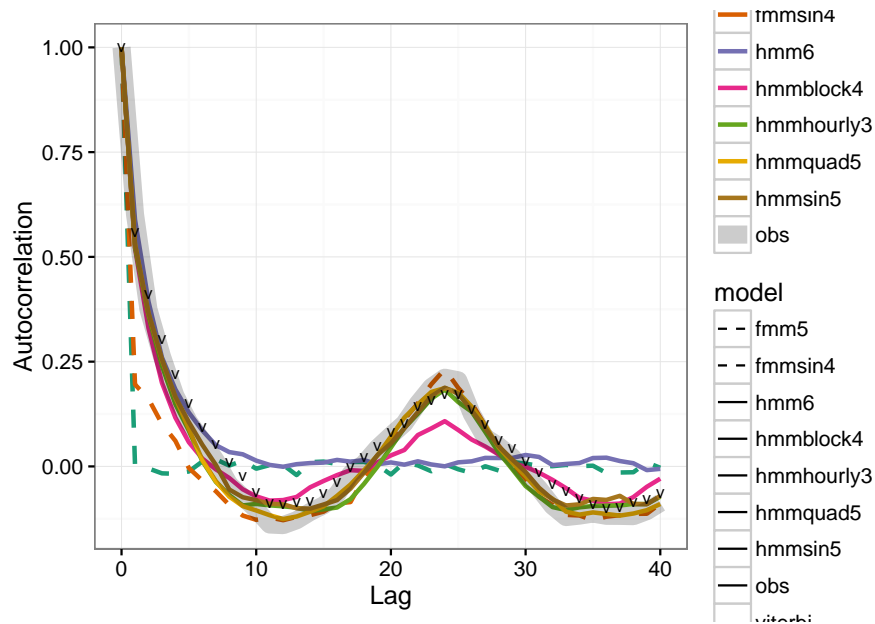


Figure 5: ACF fix similarly to Fig 3

autocorrelation beyond a 7-hour lag. Temporally homogeneous FMM, by definition, produces neither short- nor long-term autocorrelation. FMM without temporal heterogeneity, although it captures the diurnal pattern well, underpredicts the degree of short-term autocorrelation.

The hardest problem with multiple latent states is interpreting them biologically. We have no way of knowing what panthers are actually thinking (it is certainly more complex than being in one of a small number of discrete latent states); we don't know the "true" number of latent states, nor are we able to observe them directly, although incorporating additional direct observations of behaviour (if available) can at least partially address this problem (Fryxell et al., 2008). Three distinct movement states seem biologically interpretable for Florida panthers according to van de Kerk et al. (2015): Short step length suggests resting states, intermediate step length a foraging state, and long step

in figure, maybe replace/also show HMM restricted to 3 states?

maybe do Viterbi plots for mean state occupancy by hour for HMM vs HMM/TH models? Still trying to figure out what's really going on here ...

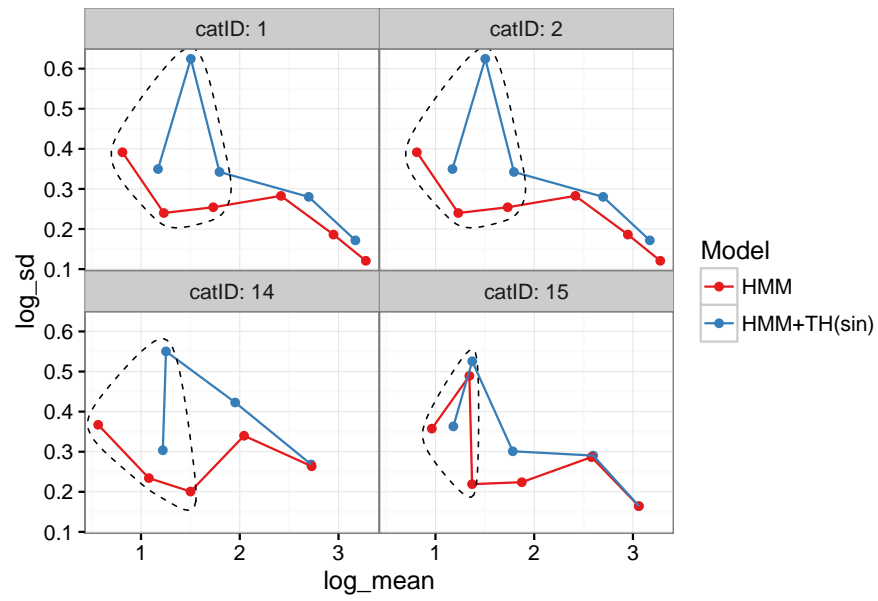


Figure 6: Step-length parameters (**todo: change scales from $\log()$ to \log -scale; need a real/more complete figure caption here; emphasize that dashes are for emphasis only; rename models (time-homog/heterog)**)

length a traveling state. Figure 6 compares the estimated parameter values for several cats (mean and standard deviation of the step length in each state) between the time-homogeneous and time-heterogeneous models. In general, the states with longer mean step lengths are relatively similar between model classes. 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).

do we need this?

4 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.

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 tran-

sitions 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.

BMB: I think we want one (or two?) paragraph(s) here that revisit(s) the deeper statistical issues of

1. that different approaches are needed for prediction vs. state identification. If we're aiming for state identification it might be OK to pre-specify the number of states and, *post hoc*, match (Viterbi-based) state estimates with environmental variation (à la Fryxell et al.). If we have specific environmental covariates in mind, though, it might be better to fit a covariate-dependent model so that we can explicitly model the switching process ... If we want to do prediction, we have to include the covariates, because Viterbi can only estimate state occupancy if observed movements are available ...
2. prediction (of movements) vs estimation of number of states also represent two different questions; (Celeux et al., 2006) (I haven't read it!) and Celeux presentation discuss this; BIC is good for prediction, but not necessarily for estimation of number of states. We don't need to go into detail, but should point in this direction. Also perhaps hint a bit more above where we reference (Dean et al., 2013).

Incorporating temporal heterogeneity 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 prin-

ciple 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 (Potts et al., 2014) 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.

(van de Kerk 2015)	Table 1: Cat-ID (IR@UF)	Number of Observations
130	1	10286
131	2	9458
48	14	14645
94	15	10250

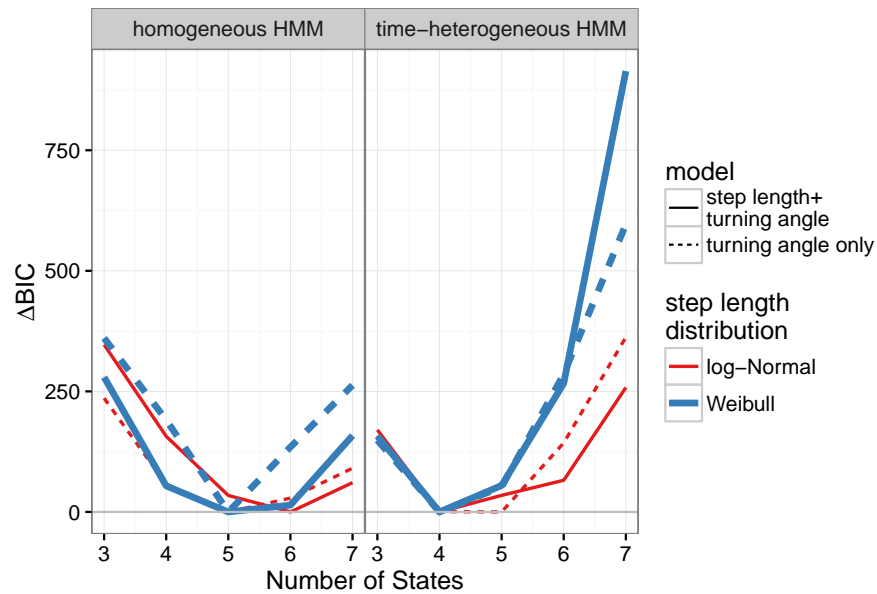


Figure 7: Change in BIC across models, cat #2. ΔBIC is scaled individually per model, so that $\Delta\text{BIC} = 0$ represents the BIC-optimal number of states for a particular model.

5 Supplementary material

```
## Error in layout_base(data, vars, drop = drop): At least one layer
must contain all variables used for facetting
```

does it make any sense to compress figs 9 etc. into a single plot (via facetgrid or multi-lines)? Same idea for all repeated plots; also, update to use percent+percent syntax

Table 2: Number of Parameters for each model type. **todo: add general formula for each model type, e.g. FMM = S (occupancies) + $2S$ (emissions: mean and std dev for each state); HMM = $S(S - 1) + 2S = S^2 + S$. Be clear about whether we're talking about restricted (constrained) or unconstrained.**

Type	Model	Number of States	Number of Parameters
FMM	FMM	3	8
FMM	FMM	4	11
FMM	FMM	5	14
FMM	FMM	6	17
FMM	FMM	7	20
FMM + TH	FMM + TH (Sin)	3	12
FMM + TH	FMM + TH (Sin)	4	17
FMM + TH	FMM + TH (Sin)	5	22
FMM + TH	FMM + TH (Sin)	6	27
FMM + TH	FMM + TH (Sin)	7	32
HMM	HMM	3	14
HMM	HMM	4	23
HMM	HMM	5	24
HMM	HMM	6	47
HMM	HMM	7	62
HMM + TH	HMM + TH (Block)	3	26
HMM + TH	HMM + TH (Block)	4	47
HMM + TH	HMM + TH (Block)	5	74
HMM + TH	HMM + TH (Block)	6	107
HMM + TH	HMM + TH (Block)	7	146
HMM + TH	HMM + TH (Quad)	3	26
HMM + TH	HMM + TH (Quad)	4	47
HMM + TH	HMM + TH (Quad)	5	74
HMM + TH	HMM + TH (Quad)	6	107
HMM + TH	HMM + TH (Quad)	7	146
HMM + TH	HMM + TH (Sin)	3	26
HMM + TH	HMM + TH (Sin)	4	47
HMM + TH	HMM + TH (Sin)	5	74
HMM + TH	HMM + TH (Sin)	6	107
HMM + TH	HMM + TH (Sin)	7	146
HMM + TH	HMM + TH (Hourly)	3	152
HMM + TH	HMM + TH (Hourly)	4	299

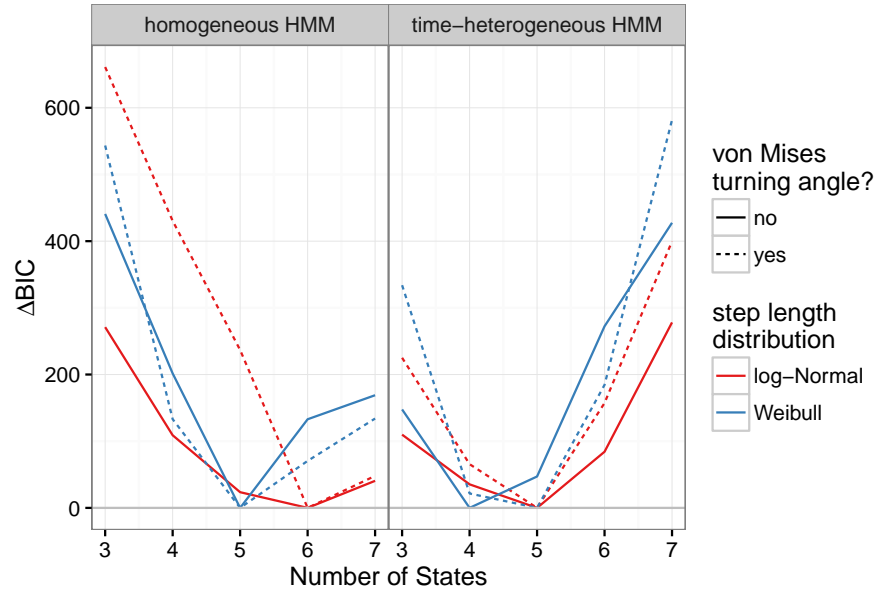


Figure 8: Change in BIC across models, cat #15. ΔBIC is scaled individually per model, so that $\Delta\text{BIC} = 0$ represents the BIC-optimal number of states for a particular model.

Figure 9: Overall adjusted BIC comparisons between different HMM models

Figure 10: Overall Adjusted BIC Comparisons between different HMM models

```
## Error in layout_base(data, vars, drop = drop): At least one layer  
must contain all variables used for facetting
```

```
## Error in layout_base(data, vars, drop = drop): At least one layer  
must contain all variables used for facetting
```

References

- Burnham, K. P. and D. R. Anderson (1998). *Model Selection and Inference: A Practical Information-Theoretic Approach*. New York: Springer.
- Celeux, G., F. Forbes, C. P. Robert, and D. M. Titterton (2006). Deviance information criteria for missing data models. *Bayesian Analysis* 1(4), 651–674.
- Dean, B., R. Freeman, H. Kirk, K. Leonard, R. A. Phillips, C. M. Perrins, and T. Guilford (2012). Behavioural mapping of a pelagic seabird: combining

Figure 11: Overall Adjusted BIC Comparisons between different HMM models

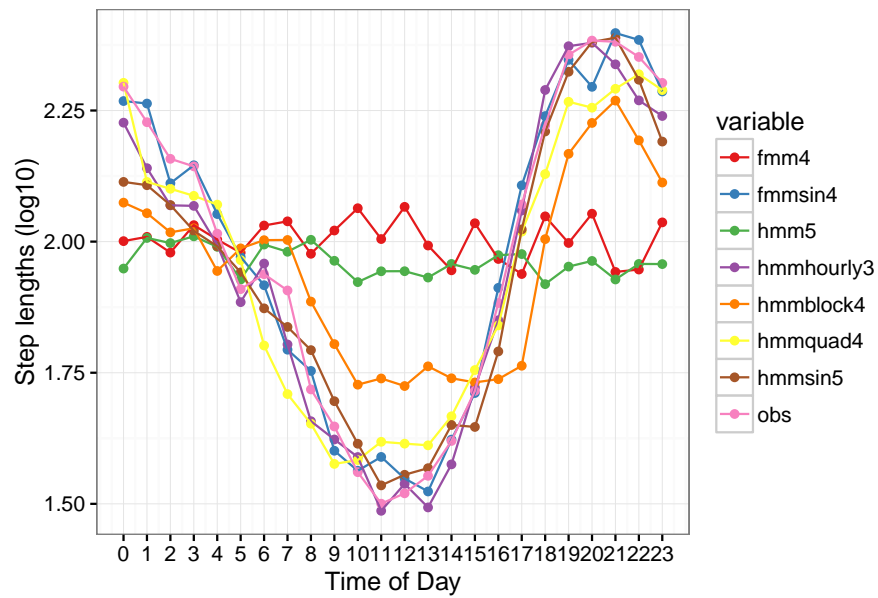


Figure 12: Average step length by time of day

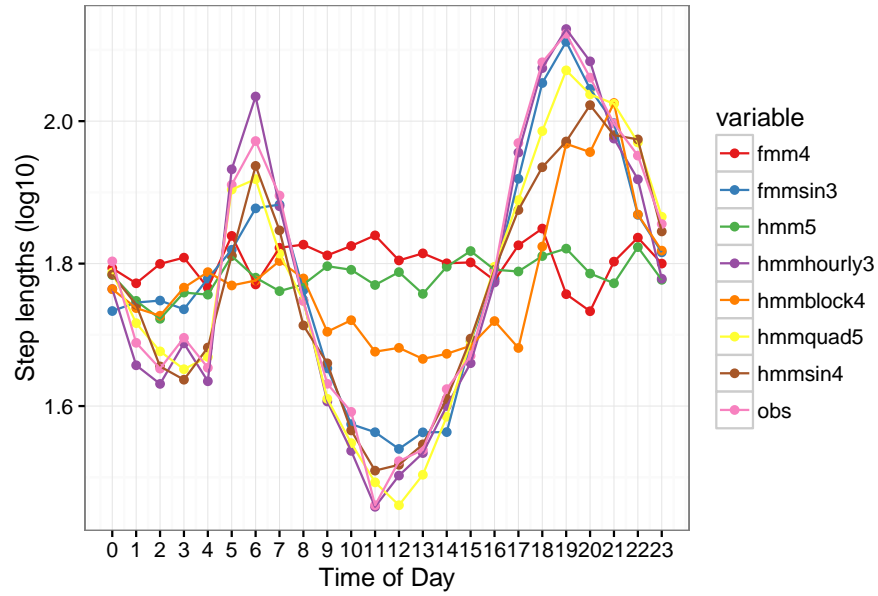


Figure 13: Average step length by time of day

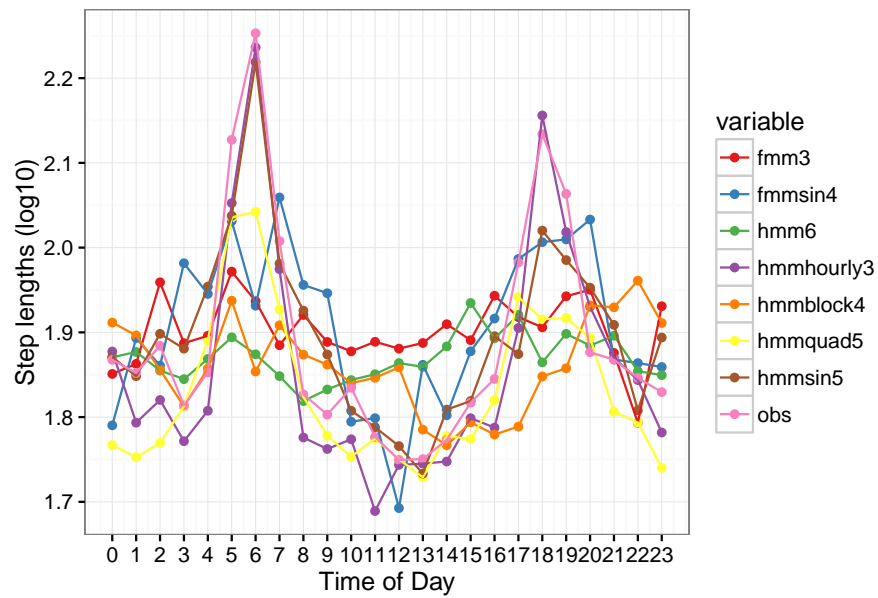


Figure 14: Average step length by time of day

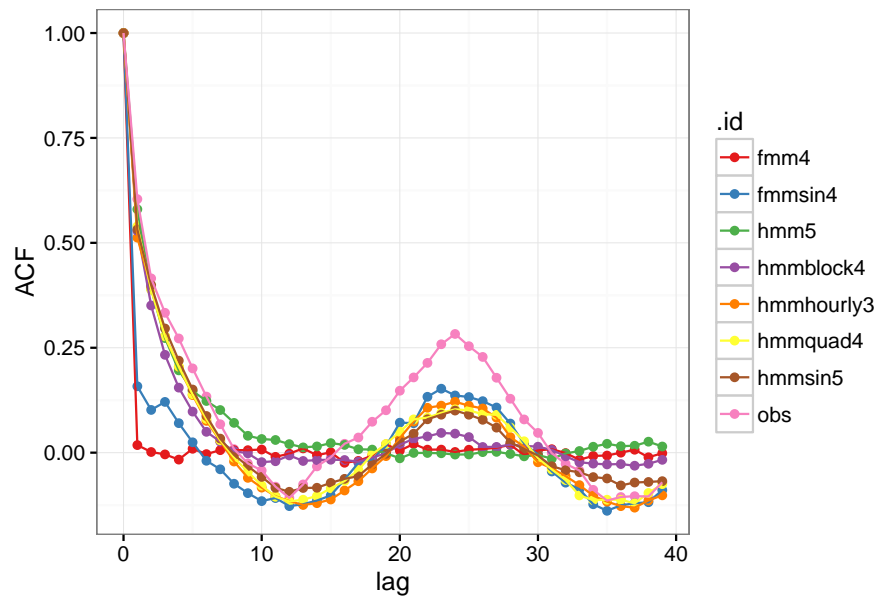


Figure 15: ACF

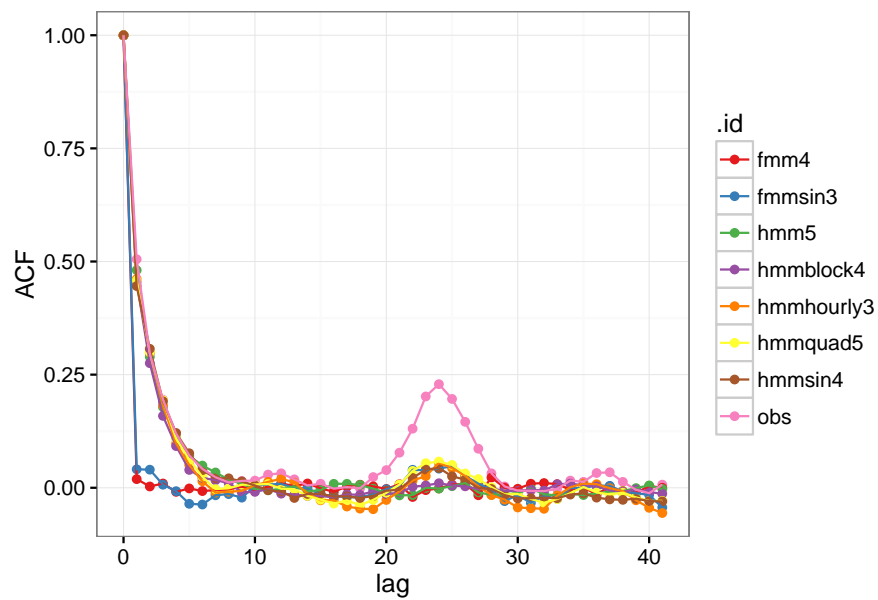


Figure 16: ACF

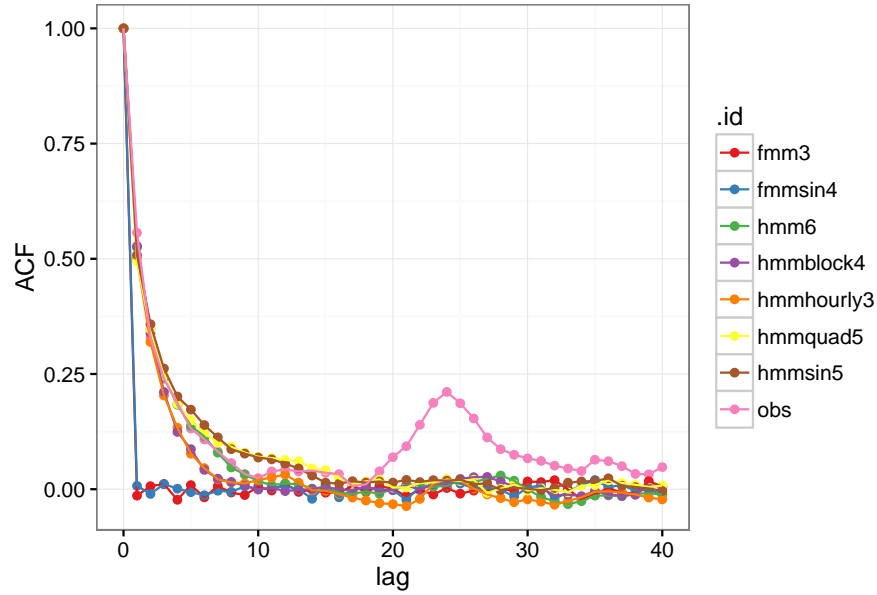


Figure 17: ACF

multiple sensors and a hidden Markov model reveals the distribution of at-sea behaviour. *Journal of the Royal Society Interface*, rsif20120570.

Dean, B., R. Freeman, H. Kirk, K. Leonard, R. A. Phillips, C. M. Perrins, and T. Guilford (2013, January). Behavioural mapping of a pelagic seabird: combining multiple sensors and a hidden Markov model reveals the distribution of at-sea behaviour. *Journal of The Royal Society Interface* 10(78), 20120570.

Firle, S., R. Bommarco, B. Ekbom, and M. Natiello (1998, September). The influence of movement and resting behavior on the range of three carabid beetles. *Ecology* 79(6), 2113–2122.

Franke, A., T. Caelli, G. Kuzyk, and R. J. Hudson (2006, August). Prediction of wolf (*Canis lupus*) kill-sites using hidden Markov models. *Ecological Modelling* 197(1–2), 237–246.

Fryxell, J. M., M. Hazell, L. Börger, B. D. Dalziel, D. T. Haydon, J. M. Morales,

- T. McIntosh, and R. C. Rosatte (2008, December). Multiple movement modes by large herbivores at multiple spatiotemporal scales. *Proceedings of the National Academy of Sciences* 105(49), 19114–19119.
- Gurarie, E., R. D. Andrews, and K. L. Laidre (2009). A novel method for identifying behavioural changes in animal movement data. *Ecology Letters* 12(5), 395–408.
- Langrock, R., R. King, J. Matthiopoulos, L. Thomas, D. Fortin, and J. M. Morales (2012, June). Flexible and practical modeling of animal telemetry data: hidden Markov models and extensions. *Ecology* 93(11), 2336–2342.
- McKellar, A. E., R. Langrock, J. R. Walters, and D. C. Kesler (2014, September). Using mixed hidden Markov models to examine behavioral states in a cooperatively breeding bird. *Behavioral Ecology*, aru171.
- McKenzie, H. W., M. A. Lewis, and E. H. Merrill (2009). First passage time analysis of animal movement and insights into the functional response. *Bulletin of Mathematical Biology* 71(1), 107–129.
- Morales, J. M., D. T. Haydon, J. Frair, K. E. Holsinger, and J. M. Fryxell (2004). Extracting more out of relocation data: building movement models as mixtures of random walks. *Ecology* 85(9), 2436–2445.
- Nathan, R., W. M. Getz, E. Revilla, M. Holyoak, R. Kadmon, D. Saltz, and P. E. Smouse (2008, December). A movement ecology paradigm for unifying organismal movement research. *Proceedings of the National Academy of Sciences* 105(49), 19052–19059.
- Okubo, A. (1980). Diffusion and Ecological Problems: Mathematical Models.
- Pal, S., B. Ghosh, and S. Roy (1998). 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.
- Patterson, T. A., M. Basson, M. V. Bravington, and J. S. Gunn (2009, November). Classifying movement behaviour in relation to environmental conditions using hidden Markov models. *Journal of Animal Ecology* 78(6), 1113–1123.
- Patterson, T. A., L. Thomas, C. Wilcox, O. Ovaskainen, and J. Matthiopoulos (2008). State–space models of individual animal movement. *Trends in Ecology & Evolution* 23(2), 87–94.
- Potts, J. R., M. Auger-Méthé, K. Mokross, and M. A. Lewis (2014). A generalized residual technique for analysing complex movement models using earth mover’s distance. *Methods in Ecology and Evolution* 5(10), 1012–1022.
- R Core Team (2015). *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing.
- Richards, S. A. (2005). Testing ecological theory using the information-theoretic approach: examples and cautionary results. *Ecology* 86(10), 2805–2814.
- Schliehe-Diecks, S., P. M. Kappeler, and R. Langrock (2012, April). On the application of mixed hidden Markov models to multiple behavioural time series. *Interface Focus* 2(2), 180–189.
- Tracey, J. A., J. Zhu, E. Boydston, L. Lyren, R. N. Fisher, and K. R. Crooks (2012, October). Mapping behavioral landscapes for animal movement: a finite mixture modeling approach. *Ecological Applications* 23(3), 654–669.
- Turchin, P. (1998). *Quantitative analysis of movement: measuring and modeling population redistribution in animals and plants*. Sunderland, MA, USA: Sinauer Associates.

- van de Kerk, M., D. P. Onorato, M. A. Criffield, B. M. Bolker, B. C. Augustine, S. A. McKinley, and M. K. Oli (2015). Hidden semi-Markov models reveal multiphasic movement of the endangered Florida panther. *Journal of Animal Ecology* 84(2), 576–585.
- Visser, I. and M. Speekenbrink (2010). `depmixS4`: An R package for hidden Markov models. *Journal of Statistical Software* 36(7), 1–21.
- Zhao, Q., M. Xu, and P. Fränti (2008, November). Knee Point Detection on Bayesian Information Criterion. pp. 431–438. IEEE.
- Zucchini, W. and I. L. MacDonald (2009, April). *Hidden Markov Models for Time Series: An Introduction Using R*. CRC Press.