Identifying latent behavioral states in animal movement with non-parametric Bayesian methods

Joshua A Cullen[1*], Caroline L Poli[2], Robert J Fletcher, Jr.[3], Denis Valle[1]

[1] School of Forest Resources and Conservation, University of Florida, Gainesville, FL, USA

[2] School of Natural Resources and Environment, University of Florida, Gainesville, FL, USA

[3] Department of Wildlife Ecology and Conservation, University of Florida, Gainesville, FL, USA

[*] Corresponding Author: 136 Newins-Ziegler Hall, Gainesville, FL 32611. PO BOX 110410. Email: joshcullen10@gmail.com

Running headline: Flexible Bayesian estimation of behavior

Article type: Research article

Abstract

1. Understanding animal movement often relies upon telemetry and biologging devices. These data are frequently used to estimate latent behavioral states to help understand why animals move across the landscape (or seascape). While there are a variety of methods that make behavioral inference from biotelemetry data, some features of these methods (e.g., analysis of a single data stream, use of parametric distributions) may result in the misclassification of behavioral states.

2. We address some of the limitations of segmentation and state-space models (SSMs) in our non-parametric Bayesian framework, available within the open-source R package `bayesmove`. This framework can analyze multiple data streams, which may capture complex behaviors more successfully than a single data stream. Additionally, parametric distributions are not used in our framework since they may poorly characterize the underlying true distributions. We tested our Bayesian framework using simulated trajectories and compared model performance against a representative segmentation method (behavioral change point analysis; BCPA) and one type of SSM, a hidden Markov model (HMM). We also illustrated this Bayesian framework using movements of juvenile snail kites (*Rostrhamus sociabilis*) in Florida, USA.

3. The Bayesian framework estimated breakpoints more accurately than BCPA for tracks of different lengths, albeit at a slower computational speed. Likewise, the Bayesian framework provided more accurate estimates of behavior than the HMM when simulations were generated from atypical distributions. This framework also performed up to three times faster than the HMM when run in a similar fashion. Three behavioral states were estimated from snail kite

movements, which were labeled as ‘encamped’, ‘area-restricted search’, and ‘transit’. Changes in these behaviors over time were associated with known dispersal events from the nest site, as well as movements to and from possible breeding locations.

4. Our framework estimated behavioral states with comparable or superior accuracy compared to BCPA and HMM when step lengths and turning angles of simulations were generated from atypical distributions. Since empirical data can be complex and do not necessarily conform to parametric distributions, methods (such as our Bayesian framework) that can flexibly classify latent behavioral states will become increasingly important to address questions in movement ecology.

Keywords: animal movement, Bayesian, behavioral state, biologging, movement ecology, non-parametric, segmentation, telemetry

1. Introduction

Understanding animal movement has advanced considerably in recent decades with the emergence of the field of movement ecology (Fraser et al. 2018; Joo et al. 2020), which focuses on understanding where animals go, what they are doing, and how they are influenced by their surrounding environment (Nathan et al. 2008; Seidel et al. 2018). As telemetry and biologging devices continue to increase in their battery life, data resolution, and affordability (Hussey et al. 2015; Kays et al. 2015; Wilmers et al. 2015), statistical methods that can efficiently analyze these large datasets will become ever more important (Patterson et al. 2017; Potts et al. 2018). To fully understand animal movement, it is necessary to account for behavior since space and resource use are directly linked to an animal's internal state (Nathan et al. 2008; Gurarie et al. 2016).

Because the direct observation of animal behavior over long time periods (months to years) is typically infeasible for wild organisms, latent behavioral states are often estimated to infer unobserved behavior from recorded tracks. These latent states can be estimated from a variety of data streams (i.e., time series of variables) such as step lengths, turning angles, ambient temperature, and acceleration, among others (Edelhoff et al. 2016). Estimation of latent behavioral states is often performed using segmentation methods or state-space models (SSMs). Segmentation methods partition tracks into segments by detecting shifts in the data stream(s), whereas SSMs estimate latent behavioral states per observation based on the transition probabilities among a given number of states (Edelhoff et al. 2016; Gurarie et al. 2016). While existing segmentation methods and SSMs provide fast or powerful predictive capacity (Edelhoff

et al. 2016; Patterson et al. 2017), respectively, they possess a number of limitations that can impact the inference made on behavioral states.

Despite their utility, existing segmentation methods and SSMs have several key limitations. For instance, segmentation methods commonly infer behavior using only a single data stream (Edelhoff et al. 2016; Gurarie et al. 2016), which can be problematic when underlying behaviors are complex and not well represented by a single metric alone. Unlike segmentation methods, SSMs typically estimate behavioral states by fitting the data streams to parametric probability distributions (e.g., Jonsen et al. 2005; Langrock et al. 2012; Patterson et al. 2017), such as a gamma or wrapped Cauchy distributions. The neglected structure in the data streams not captured by parametric distributions can often result in the overestimation of the true number of states by information criteria (Pohle et al. 2017). Importantly, estimating SSMs can be computationally costly: model runtime can take minutes to days depending on the type of SSM, sample size, number of estimated states, and computer hardware. Furthermore, both ML and Bayesian SSMs are often run multiple times to determine the optimal number of behavioral states (Jonsen et al. 2013; Pohle et al. 2017), which can be quite computationally intensive.

Given the limitations posed by existing methods that estimate latent behavioral states, there is a need to develop a framework that is based on as few parametric assumptions as possible while also being fast and flexible. Here, we introduce a new two-stage modeling framework that implements non-parametric Bayesian methods to: 1) segment multiple data streams into relatively homogeneous units of behaviors; and 2) subsequently determine the likely number of behavioral states using a mixed-membership method. Latent behavioral states are estimated for track segments (as opposed to observations) since this reflects our understanding that behavior is inherently autocorrelated, especially when observations are sampled at short time

intervals (Pohle et al. 2017; Potts et al. 2018). This model framework is available within the open-source R package `bayesmove` available on CRAN (Cullen and Valle 2020). Furthermore, we describe the model structure and the Bayesian sampling methods used to estimate from the posterior distribution. We also demonstrate that this model can successfully recover breakpoints and behavioral states based on simulated trajectories and compare our model's performance against a frequently used segmentation method (behavioral change point analysis; Gurarie et al. 2009) and one type of SSM, a hidden Markov model (HMM). Finally, we illustrate our novel approach on the movements of an endangered raptor species, the Everglade snail kite (*Rostrhamus sociabilis*), and interpret the results within the context of natal and breeding dispersal events.

2. Materials and Methods

2.1 Model Structure

2.1.1 Discretization of data streams

We circumvent some of the limitations associated with existing segmentation methods and SSMs by applying our new framework on discretized data streams (Figs 1a, b, c). Although data streams (e.g., step lengths and turning angles) are not typically discretized into bins, this process can accommodate distributions that may not conform to parametric density distributions and may also be more robust to extreme values. Robustness arises because bins are estimated independently of one another and extreme values lose their influence when added to the first or last bins with the rest of the data. Therefore, the discretization of data streams is expected to increase model flexibility.

2.1.2 Segmentation

We estimate the location and number of unknown breakpoints along a time series of data streams by implementing a Gibbs sampler within a reversible-jump Markov chain Monte Carlo (RJMCMC) algorithm. RJMCMC is a trans-dimensional algorithm that provides simultaneous inference on parameter values given a particular model, as well as model space (i.e., the collection of all possible models) (Green 1995; Sisson 2005). We adopt this approach to perform unsupervised segmentation on each individual trajectory. In our framework, each potential model M_k is characterized by a set of breakpoints $\{b_{1k}, \ldots, b_{Pk}\}$, where k is the model number. Each breakpoint is restricted to being an integer between 2 and T_i across all observations, where T_i is the total number of observations for individual i. Given a particular model M_k, its breakpoints define track segments. We assume that for any given track segment c:

$$x_{itj} \sim Categorical(\boldsymbol{\theta_{cj}}),$$

where x_{itj} is the bin label for individual i at time t for data stream j and $\boldsymbol{\theta_{cj}}$ is a vector of probabilities that sum to one. The use of a categorical distribution to characterize track segments (as opposed to continuous distributions) is what makes this framework non-parametric. Our prior is given by:

$$\boldsymbol{\theta_{cj}} \sim Dirichlet(\alpha),$$

where α is equal across all bins. We integrate over the latent parameter $\boldsymbol{\theta_{cj}}$ to enable the algorithm to visit multiple models and increase computational efficiency. Since there are no

longer any remaining parameters on which to assess model convergence, we do so by evaluating trace-plots of the log marginal likelihood (Denison et al. 2002). Details for the derivation of the full conditional distributions can be found in the Supporting Information.

Since only a single set of breakpoints can be used in downstream analysis and models can vary greatly in their number and position of breakpoints, we select the Maximum a Posteriori (MAP) estimate (i.e., the breakpoints of the model with the greatest log marginal likelihood). The MAP estimate is used in place of the full posterior since the number and position of these breakpoint estimates vary, which makes it difficult to incorporate breakpoint uncertainty in our subsequent model. However, the MAP estimate appears to be in good agreement with the breakpoints estimated from the entire posterior distribution as described in the Supporting Information. These breakpoints are then used to create track segments of data streams per individual that will be evaluated within the mixed-membership model for behavioral state estimation (Fig. 1d).

2.1.3 Mixed-Membership Clustering

The use of Latent Dirichlet Allocation (LDA; a mixed-membership model) in combination with Bayesian non-parametrics facilitates the estimation of a limited number of behavioral states. Unlike existing methods that assign a single behavior to observations or track segments, LDA classifies each track segment as a combination of multiple behaviors (Blei et al. 2003; Valle et al. 2014). This conceptually fits with our notion that track segments of animal movement may not be entirely comprised of a single behavior (Pohle et al. 2017; Patin et al. 2020). However, LDA can also classify track segments as belonging to a single behavior if that is supported by the data (Valle et al. 2014). We use a truncated stick-breaking (TSB) prior from

Bayesian non-parametrics to facilitate the simultaneous estimation of the optimal number of behavioral states and model parameters (Valle et al. In Review; Valle et al. 2017). This is an improvement on existing methods in the sense that our clustering model only needs to be run once, whereas SSMs are typically run multiple times with varying numbers of behavioral states to then determine the best model via information criteria (e.g., AIC or BIC).

The LDA evaluates track segments from all individuals simultaneously since we assume that there is a common set of behaviors exhibited across all individuals of the population. Although there may be some individual heterogeneity in movement patterns, the pooling of all individuals ensures that behavioral states are directly comparable and improves the inference on individuals with fewer observations (Jonsen 2016). In this model, we assume that:

$$y_{ijct} \mid z_{ijct} = k \sim Categorical(\boldsymbol{\phi_{kj}}),$$

where y_{ijct} denotes the bin for time t of track segment c from data stream j for individual i. Additionally, $z_{ijct} = k$ is the latent cluster membership associated with each of the possible behavioral states k and $\boldsymbol{\phi_{kj}}$ is a vector of probabilities for each behavior and data stream. Notice that z_{ijct} influences the distribution for y_{ijct} by determining the subscript k for the vector $\boldsymbol{\phi_{kj}}$. We assume that the latent cluster membership is given by:

$$z_{ijct} \sim Categorical(\boldsymbol{\theta_{ic}}),$$

where $\boldsymbol{\theta_{ic}}$ is a vector of probabilities of size K (i.e., the number of clusters or behaviors) that sum to one and indicates the likelihood of assigning an observation at time t of track segment c for

individual i to each behavioral state k. This formulation assumes that each observation within a particular track segment must belong to a single behavioral state, but that track segments are comprised of multiple behaviors. For our priors, we assume that:

$$\boldsymbol{\phi_{kj}} \sim Dirichlet(\alpha)$$

$$\boldsymbol{\theta_{ic}} \sim TSB(\gamma),$$

where $TSB(\gamma)$ represents the truncated stick-breaking prior. This prior is given by:

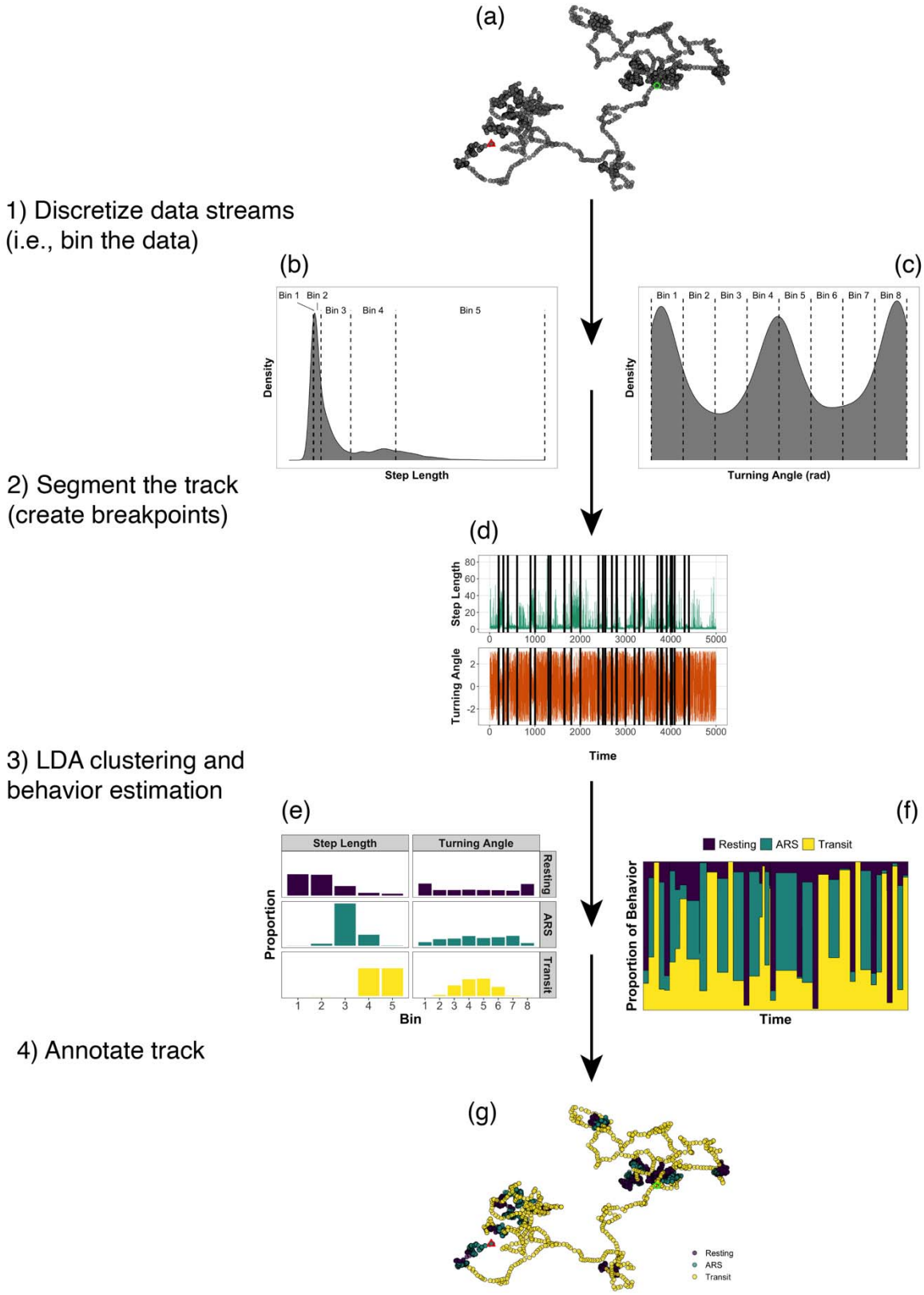
$$\theta_{ick} = V_{ick} \prod_{p=1}^{K-1} (1 - V_{icp})$$

$$V_{ick} \sim Beta(1, \gamma),$$

where $V_K = 1$ and $\gamma > 0$. By setting $0 < \gamma < 1$, we can shrink θ_{ick} to approximately zero for large k. As a result, fewer and fewer observations will be assigned to behaviors with larger and larger k, enabling the model to identify the optimal number of behavioral states.

This LDA model is fit using a Gibbs sampler and a complete description of the full conditional distribution can be found in the Supporting Information. Similar to the segmentation model, convergence was assessed by inspecting trace-plots of the log-likelihood. The posterior mean for all $\boldsymbol{\theta_{ic}}$ was then used to identify the optimal number of behaviors. The estimated state-dependent distributions for each data stream (from $\boldsymbol{\phi_{kj}}$) were evaluated and used to corroborate the findings based on the posterior means from all $\boldsymbol{\theta_{ic}}$'s by determining whether the distributions were biologically relevant (Figs 1e, f, g). This combination of results provides a straightforward

approach to selecting the optimal number of behavioral states. A list of the primary functions to analyze data using the `bayesmove` R package is included in Table S1.


(a)
1) Discretize data streams
(i.e., bin the data)
(b)
Bin 1
Bin 2
Bin 3
Bin 4
Bin 5
Density
Step Length
(c)
Bin 1
Bin 2
Bin 3
Bin 4
Bin 5
Bin 6
Bin 7
Bin 8
Density
Turning Angle (rad)
2) Segment the track
(create breakpoints)
(d)
Step Length
Turning Angle
Time
3) LDA clustering and
behavior estimation
(e)
Step Length
Turning Angle
Resting
ARS
Transit
Proportion
Bin
(f)
Resting
ARS
Transit
Proportion of Behavior
Time
4) Annotate track
(g)
Resting
ARS
Transit

219

Figure 1 *General workflow to analyze animal trajectories from telemetry data. Steps from this analysis include: (a) starting with a track of coordinates that are used to calculate step lengths and turning angles, (b) discretizing step lengths and (c) turning angles into sets of bins, (d) performing segmentation on the joint time series of step lengths and turning angles, (e) cluster the track segments into behavioral states by Latent Dirichlet Allocation (LDA) and (f) evaluate time series of behavior proportions for each individual, and (g) visualize the annotated tracks by displaying the dominant behavior of each time series. Note that continuous time series are only displayed with breakpoints (black lines) in (d) to improve interpretability of segments.*

2.2 Simulation Study

We assessed the performance of our model compared to other methods via simulations. We first evaluated the ability of our track segmentation method to detect true breakpoints and compared its results to those obtained by a representative segmentation method (behavioral change point analysis, BCPA; Gurarie et al. 2009). We then evaluated the ability of our clustering method to estimate the true number of behavioral states and to properly assign behavior proportions to track segments. For this component, we compared the results of our model to those obtained by a HMM (McClintock and Michelot 2018).

2.2.1 Generating a simulated trajectory

We generated multiple three-state trajectories from a correlated random walk at regular time intervals, where five tracks were simulated at each of four durations (1000, 5000, 10000, 50000 observations), resulting in a total of 20 tracks. Each track was comprised of 10, 50, 100, or 500 segments that each included 100 observations. Each of these segments included a dominant behavioral state (80% of observations), which were randomly assigned to each segment. This resulted in some instances where consecutive segments shared the same dominant state. The three behavioral states were parameterized to represent (1) little to no movement (‘encamped’),

(2) slow and tortuous movement ('area-restricted search' or ARS), as well as (3) fast and directed movement ('transit'). In our simulations, step lengths for each behavior were drawn from a truncated normal distribution and turning angles were drawn from a mixture of beta, uniform, and truncated normal distributions for the 'encamped', 'ARS', and 'transit' states, respectively (Fig. 2). The goal here was to determine how well HMMs could estimate the latent behavioral states when the assumed parametric distributions for step lengths and turning angles were not identical to those used to generate the data.

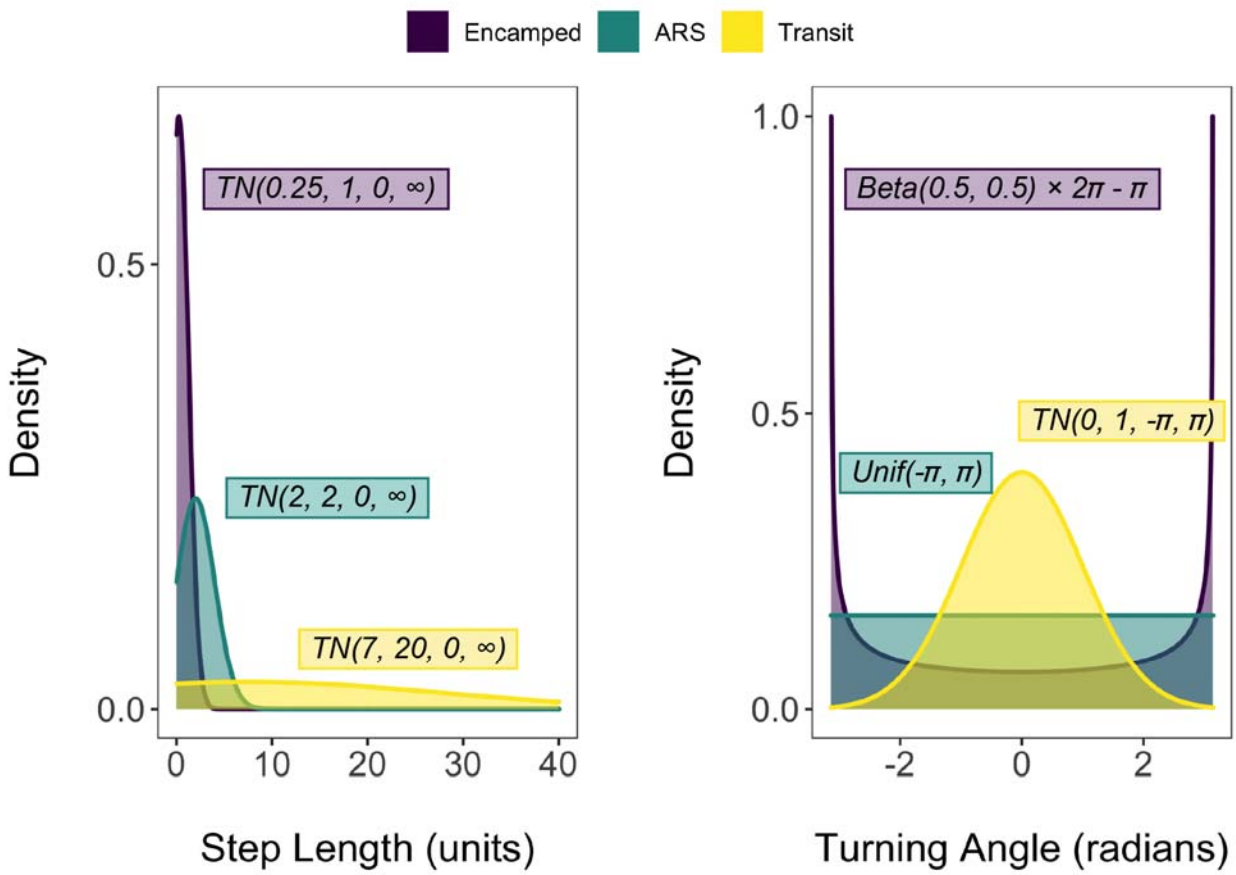


Figure 2 Distributions and parameters used to generate step lengths and turning angles for each simulated state. Step lengths are generated from a truncated normal (TN) and turning angles are generated from beta, uniform, and truncated normal distributions. $TN(a,b,c,d)$ denotes a truncated normal distribution with mean a, standard deviation b, lower bound c, and upper bound d.

2.2.2 Implementation of Bayesian framework

Step lengths and turning angles were the data streams used to make inference on latent behavioral states. Step lengths were separated into five bins using the 25th, 50th, 75th, 90th, and 100th quantiles as upper limits. These quantiles were used to discretize step lengths since this variable was highly right-skewed and the use of equal bin widths would result in very unbalanced observations across bins. Turning angles were discretized into eight bins from $-\pi$ to π using equal bin widths ($\frac{\pi}{4}$) since the distribution of this variable was relatively balanced.

Selecting the number of bins and the binning method is relatively subjective and therefore it is important that prior biological reasoning be used to inform these decisions. In particular, the number of bins should be sufficient to characterize the shape of the density distribution and there should be biologically meaningful differences in the movement pattern represented by the different bins. These assumptions during the discretization process are not unlike assumptions made for HMMs when selecting probability distributions to fit data streams, but require practitioners to make more decisions up front.

Each simulated track was analyzed by our segmentation model using 40,000 MCMC iterations and a vague prior where hyperparameter α was set to 1. We then assessed how well our model identified the true breakpoints of the behavior time series for all simulations, where a threshold of $\pm$ 10 observations was used to distinguish an accurate from an inaccurate estimated breakpoint. If no estimated breakpoints were within $\pm$ 30 observations from a true breakpoint, then the model was judged to have missed that breakpoint. While subjective, these thresholds were selected to account for minor deviations of the model estimates from the true breakpoints. The use of other thresholds resulted in the same relative pattern of accuracy measures when comparing among simulations of different lengths (JA Cullen, unpublished data).

Estimated track segments were used as input for the LDA model, which was run using 1000 MCMC iterations, a burn-in of 500 iterations, and vague priors where hyperparameters were set to $\gamma = 0.1$ and $\alpha = 0.1$. We tested for up to a maximum of seven behavioral states since this was expected to likely include the true number of states, which would be estimated via the truncated stick-breaking prior of the LDA. Trace-plots of the log-likelihood indicated that the model had converged for each simulation. The true number of behaviors was estimated by calculating the arithmetic mean of behavior proportions across all track segments and selecting the clusters (states) that represented at least 90% of all observations on average. This was supported by an inspection of step length and turning angle distributions so that we only selected common behaviors that were also biologically interpretable. Since the LDA is a mixed-membership model that treats track segments as a combination of behavioral states, proportions of each behavior were estimated per track segment. Accuracy of behavioral state estimates across track segments were evaluated by two methods: 1) we calculated the proportion of observations for which the dominant behavior of each track segment was accurately classified, and 2) we calculated the root mean square error (RMSE) of the estimated behavior proportions compared to the true behavior proportions over all states and track segments.

2.2.3 Method comparison

We compared the performance of our model on the simulated trajectories against BCPA and HMM. All models were run using a 2.6 GHz i7 CPU with 16 GB RAM. The BCPA model performed segmentation on the combination of step lengths and turning angles as persistence velocity ($Vcos(\theta)$) using the R package `bcpa` v1.1 (Gurarie 2014). Parameters for this analysis

were tuned to provide a close approximation of the true number and location of simulated breakpoints with window size set to 80, sensitivity set to 2, and clusterwidth set to 30.

A discrete-time HMM was also fitted to each of the simulated trajectories using the R package `momentuHMM` v1.5 (McClintock and Michelot 2018). Step lengths were modeled using a gamma distribution and turning angles were assumed to arise from a wrapped Cauchy distribution. The HMMs for each simulation were run using a range of 2 to 4 possible behavioral states (K) and each K-state model was run 30 times using different starting values to increase the chance of finding the global (as opposed to local) maximum of the likelihood. The selection of "good" starting values is critical since it can affect computation time and the ability of the model to identify the global maximum of the likelihood (Michelot et al. 2016; McClintock and Michelot 2018). The optimal number of states was selected using a combination of AIC and BIC, where the model with the lowest value was considered to be optimal. However, if the difference in AIC or BIC (ΔAIC/BIC) of the next best model was < 10 (Burnham and Anderson 2002), the more parsimonious model was chosen. Behavior classification accuracy was assessed at the segment-level so that results were directly comparable with the Bayesian model. This was achieved by using true breakpoints to segment the time series of states estimated by the HMM and then calculating the proportion of these behaviors within each track segment. Additionally, the resulting state-dependent distributions of step lengths and turning angles were discretized using the bin limits defined for the Bayesian model to compare the accuracy of distribution shapes. Accuracy was measured by RMSE across bins of all states and data streams per simulation.

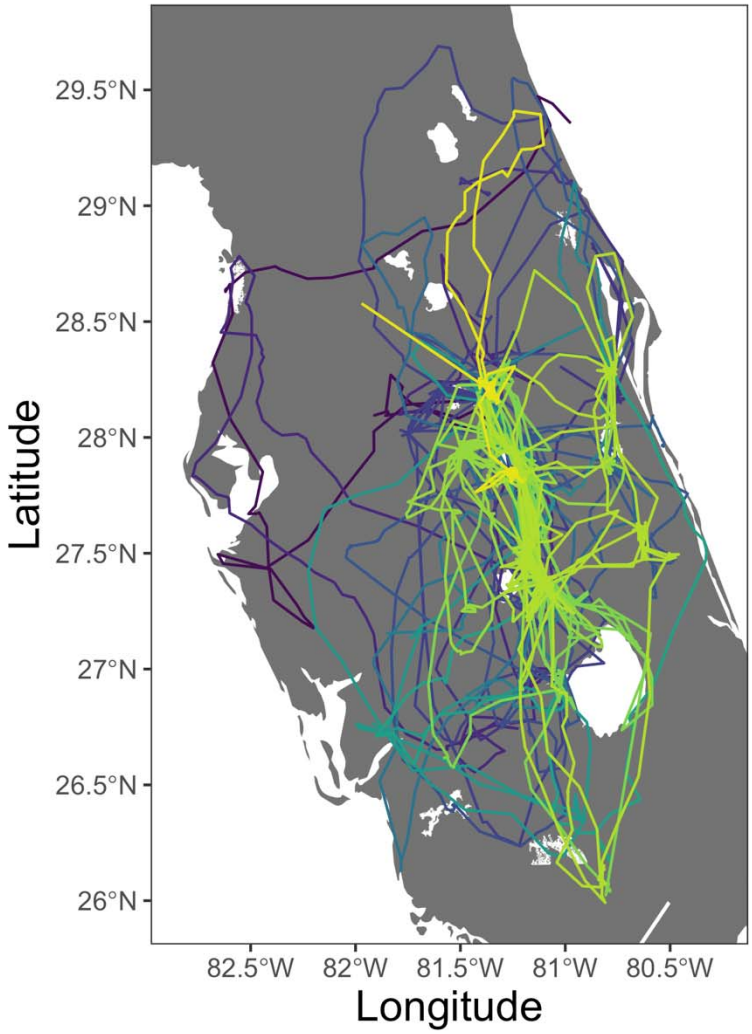


Figure 3 *Tracks from 27 GPS-tagged snail kites in central Florida (2018-2019), where colors indicate different individuals.*

2.3 Snail kite case study

As part of a larger investigation on the effects of wetland management on wildlife, solar-powered GPS-GSM transmitters (Ecotone Telemetry) were attached to juvenile snail kites (n=27) prior to fledging at Lakes Tohopekaliga, East Tohopekaliga, and Kissimmee in central Florida during 2018 and 2019. Subsequent movement of each individual over months-years

resulted in a total of 40,720 observations (Fig. 3). Locations were collected once per hour only during daylight at an accuracy of ±30 m. As a result of the programmed duty cycle and time periods where GPS tags failed to transmit data, track time intervals were irregular. To ensure comparable step lengths and turning angles, we filtered our data to the most common time interval (i.e., 1 h). We chose to omit all other observations since imputation procedures for long time gaps would increase the number of artificial data and the use of linear interpolation would artificially inflate the number of turning angles at zero radians. This filtering process caused the exclusion of a single individual that did not have any observations at a 1 h time interval, resulting in 26 individuals retained for all subsequent analyses.

Step lengths and turning angles were used to estimate latent behavioral states. As was performed on the simulated tracks, step lengths were discretized into five bins using the 25th, 50th, 75th, 90th, and 100th quantiles as upper limits. This resulted in bin limits at 0.00, 0.03, 0.09, 0.32, 1.63, and 72.56 km. Turning angles were discretized into eight bins from $-\pi$ to π using equal bin widths of $\frac{\pi}{4}$, resulting in bin limits at $-\pi, -\frac{3\pi}{4}, -\frac{\pi}{2}, -\frac{\pi}{4}, 0, \frac{\pi}{4}, \frac{\pi}{2}, \frac{3\pi}{4}, \pi$.

Step lengths and turning angles for each of the 26 snail kites were segmented by the RJMCMC algorithm using 40,000 iterations and with hyperparameter $\alpha = 1$. The MAP estimates of breakpoints for each snail kite were used to define track segments per individual. Subsequently, track segments were analyzed across all individuals via LDA to estimate the optimal number of states, to define state-dependent distributions, and to estimate the proportion of each state that characterized each track segment. This was performed using 1000 MCMC iterations with a burn-in of 500 iterations and vague priors were used with hyperparameters set at $\gamma = 0.1$ and $\alpha = 0.1$ for a maximum possible number of seven states. Proportions of behavioral states were evaluated over time in relation to emigration from natal sites and peak breeding

season of snail kites in Florida (March 1 – June 30) to discern any patterns associated with these events (Reichert et al. 2020).

3. Results

3.1 Method comparison

The Bayesian segmentation model successfully recovered breakpoints from the simulations and outperformed BCPA. However, the Bayesian segmentation model took longer to run (2 to 120 min) compared to the BCPA model (0.25 to 21 min) (Fig. 4a). The Bayesian model estimated 55.3%, 83.2%, 83.1%, and 80.4% of breakpoints accurately on average for tracks of 1000, 5000, 10000, and 50000 observations in length, respectively (Figs 4b, 5a). By comparison, BCPA estimated only 32.9%, 30.3%, 31.2%, and 18.7% of breakpoints accurately on average for tracks of increasing length (Fig. 4b). Additionally, BCPA missed 3× more of the 2034 true breakpoints than the Bayesian model.

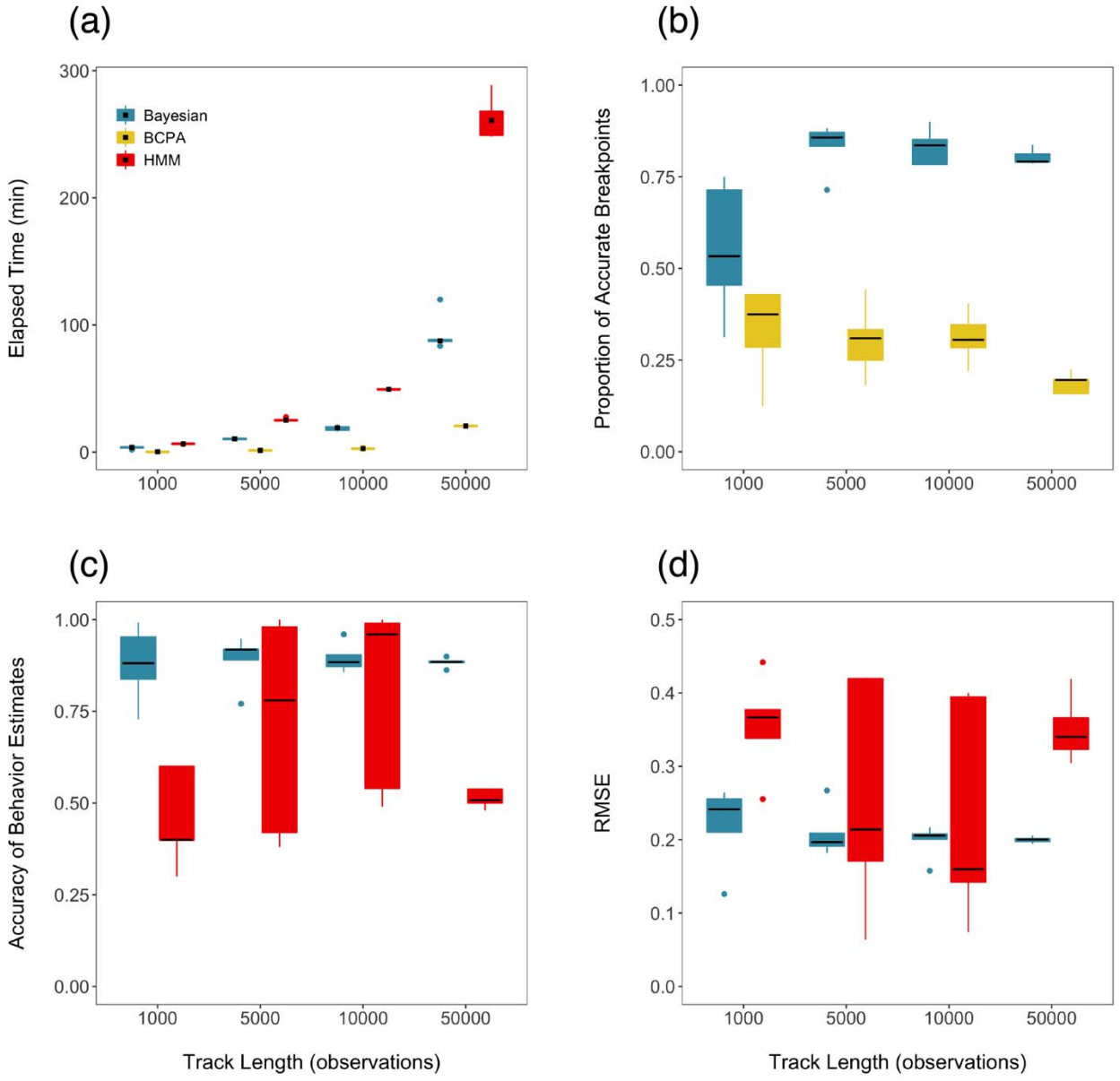


Figure 4 *Results are shown to compare the performance of different methods, where black squares or lines indicate the median for each boxplot. (a) The elapsed time to analyze each of the simulations is shown for the different methods, where the measure for the Bayesian method reflects the sum of elapsed times from both the segmentation and LDA models. (b) Accuracy of breakpoint estimates are compared between the BCPA and Bayesian models. (c) Accuracy of the estimates for the dominant behavior of each track segment is shown for the Bayesian method and HMM. (d) Accuracy of behavior proportion estimates per track segment are compared between the Bayesian method and HMM.*

When compared to the HMM, the Bayesian model correctly determined the number of true states more frequently and exhibited greater computational efficiency. The Bayesian LDA model took 2 – 23 s to run, highlighting the computational efficiency of this particular model. When added to the duration of the segmentation model, the Bayesian framework ran faster than the HMM at all track lengths despite HMMs being fitted with only two to four states (Fig. 4a). The LDA model also correctly suggested three states as optimal for 18 of the 20 simulations (Fig. 5b, Table S2), which was slightly greater than the 17 and 16 simulations suggested by AIC and BIC, respectively, for the HMM. To enable direct comparisons between the Bayesian model and HMM, we assumed three states were optimal for all 20 simulations when calculating model accuracy.

The Bayesian model demonstrated equivalent or superior accuracy in behavioral state estimation compared to the HMM, particularly for the shortest and longest tracks. Mean accuracy of the three-state HMM for classifying the dominant behavioral state within each segment was 46.0%, 71.2%, 79.6%, and 51.3% for tracks of 1000, 5000, 10000, and 50000 observations, respectively (Fig. 4c). By comparison, mean accuracy of the Bayesian model was 87.8%, 88.9%, 89.5%, and 88.3% for tracks of increasing length, which was 1.1-1.9× greater than the HMM on average (Fig. 4c). Additionally, accuracy measures displayed lower variability in the Bayesian model, which highlights the increased stability of this framework. Similar to the pattern found for estimates of dominant behavioral states, accuracy of behavioral state proportions were 1.2-1.8× higher in the Bayesian model as denoted by low RMSE (Figs 4d, 5c).

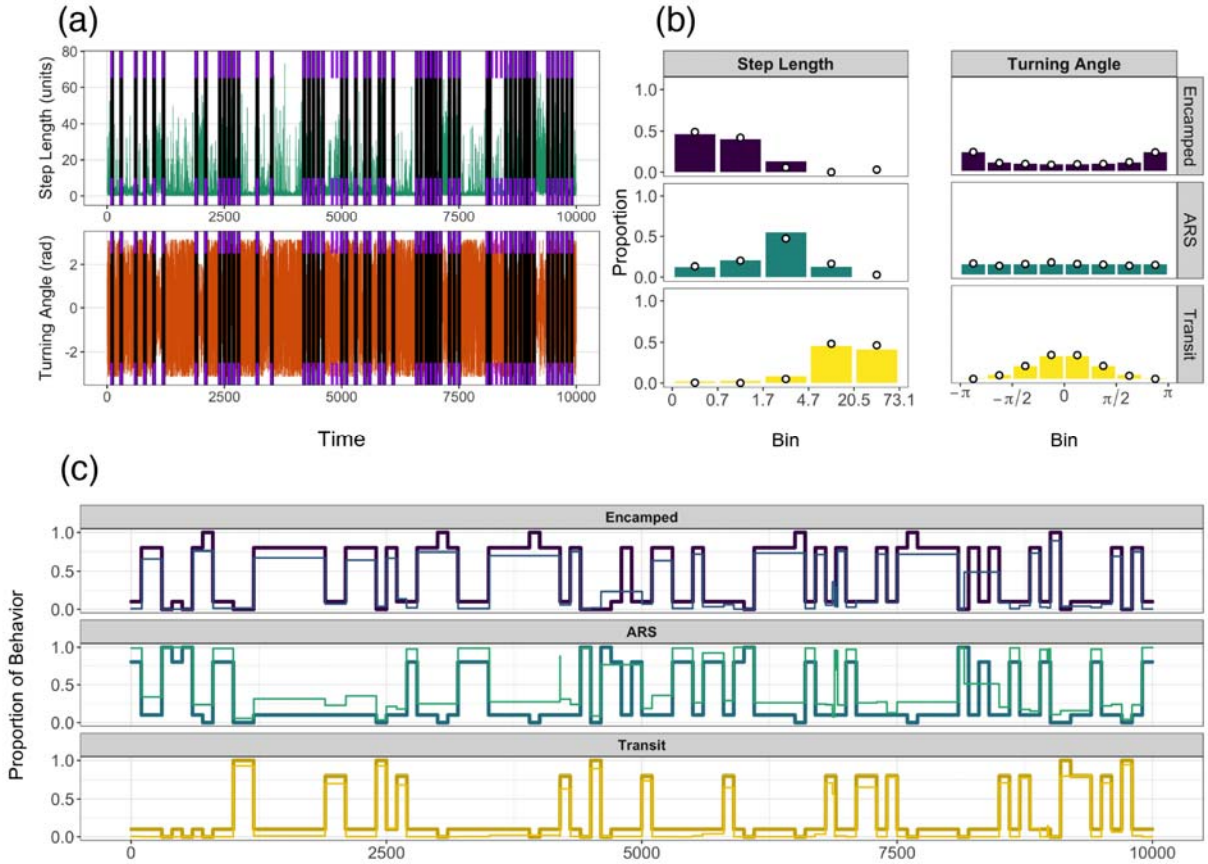


Figure 5 *Evaluation of model performance for a single simulation, as an example, for (a) breakpoint estimation, (b) determination of the shape of behavioral states, and (c) estimating behavior proportions over time. (a) A time series of true (purple tick marks) and estimated breakpoints (black lines) are shown for the segmentation of both data streams. (b) Discretized distributions are shown for each data stream per behavioral state where three states were determined to be optimal. Bars denote true proportions for each of the bins whereas points indicate model estimates. (c) Time series of proportions for each behavior are shown; thick, dark lines indicate true behavior proportions whereas thin, light lines indicate estimated proportions of behaviors.*

The distribution shapes of step lengths and turning angles were also more accurate in the Bayesian model compared to the HMM. When viewed as continuous distributions, it is clear that step lengths of the 'ARS' state and turning angles of the 'encamped' state were not well

estimated by the HMM regardless of simulated track length (Fig. 6). While comparable at the shortest track length, the Bayesian model characterized the state-dependent distributions of both step lengths and turning angles more accurately (as measured by RMSE) for the remaining tracks (Fig. S1).

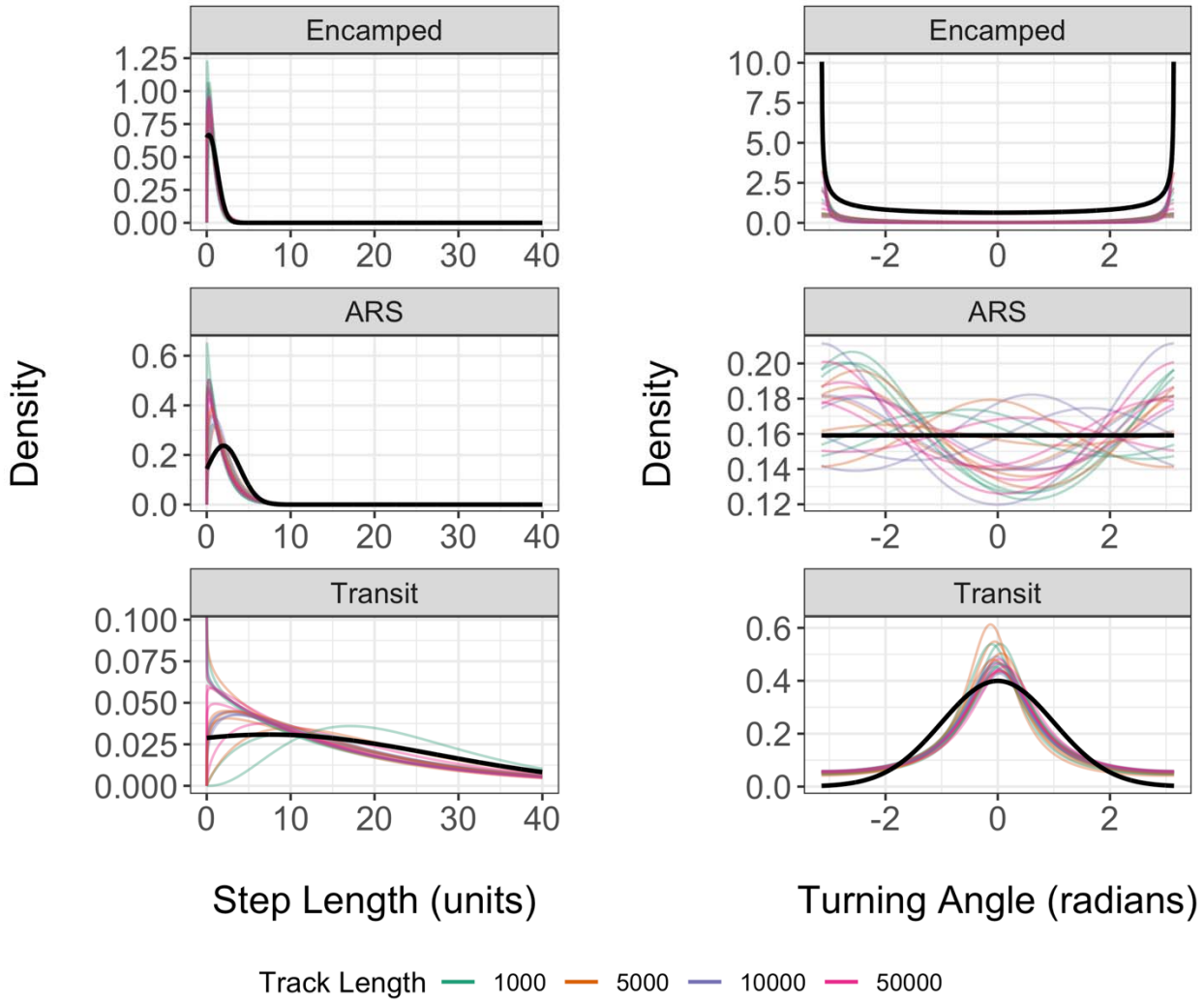


Figure 6 *Distributions of step lengths and turning angles returned by the HMMs (thin lines) are compared against the true underlying distributions (thick, black line). The HMMs appeared to have trouble fitting step lengths for the 'ARS' state as well as for the turning angles of the 'encamped' state. Note that the scale of the y-axis varies by plot.*

3.2 Snail kite analysis

The segmentation of 26 snail kite trajectories using the Bayesian model took approximately 25 min to run and it estimated 1 to 51 breakpoints for these individuals. Breakpoints were then used to define 439 track segments from all 26 individuals (Fig. 7a). These segments were analyzed by LDA, which took approximately 16 s to run. Based upon estimates from the posterior distribution, it appeared that there were likely three behavioral states, which comprised 94.1% of all behavior assignments on average (Fig. 7b). To ensure that these three states were biologically interpretable, distributions of step lengths and turning angles were also evaluated (Fig. 7c). The distributions showed: 1) a slow and tortuous behavior; 2) a tortuous behavior with intermediate speed; and 3) a fast and directed behavior. For this reason, these behaviors were labelled ‘encamped, ‘ARS’, and ‘transit’, respectively.

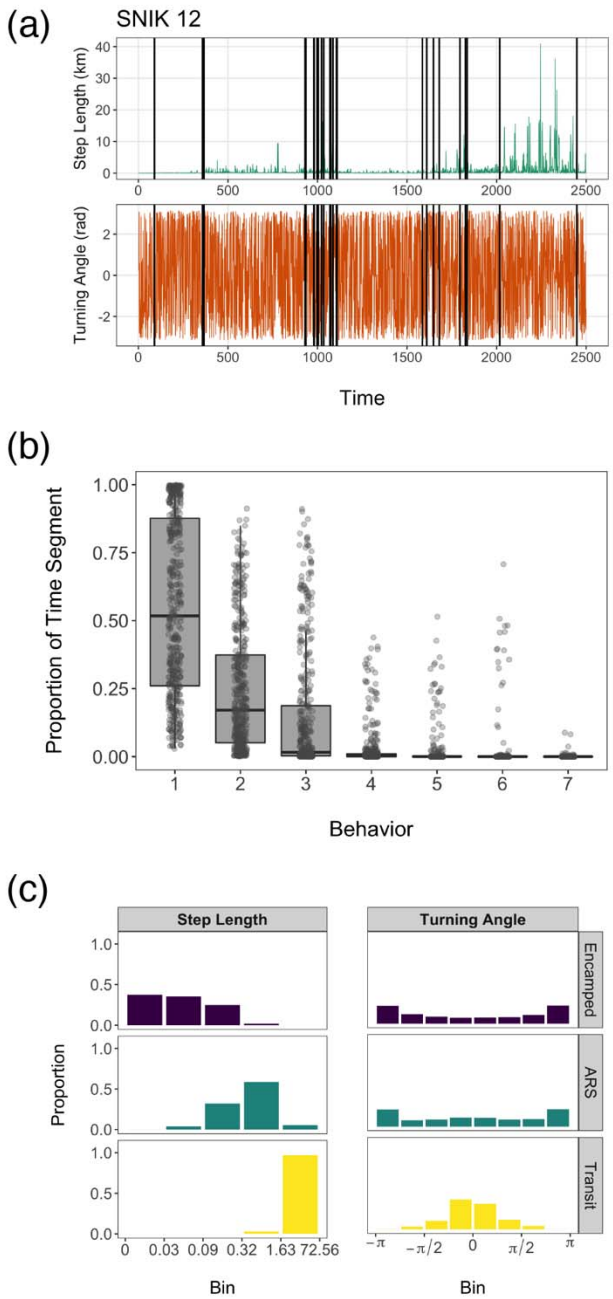
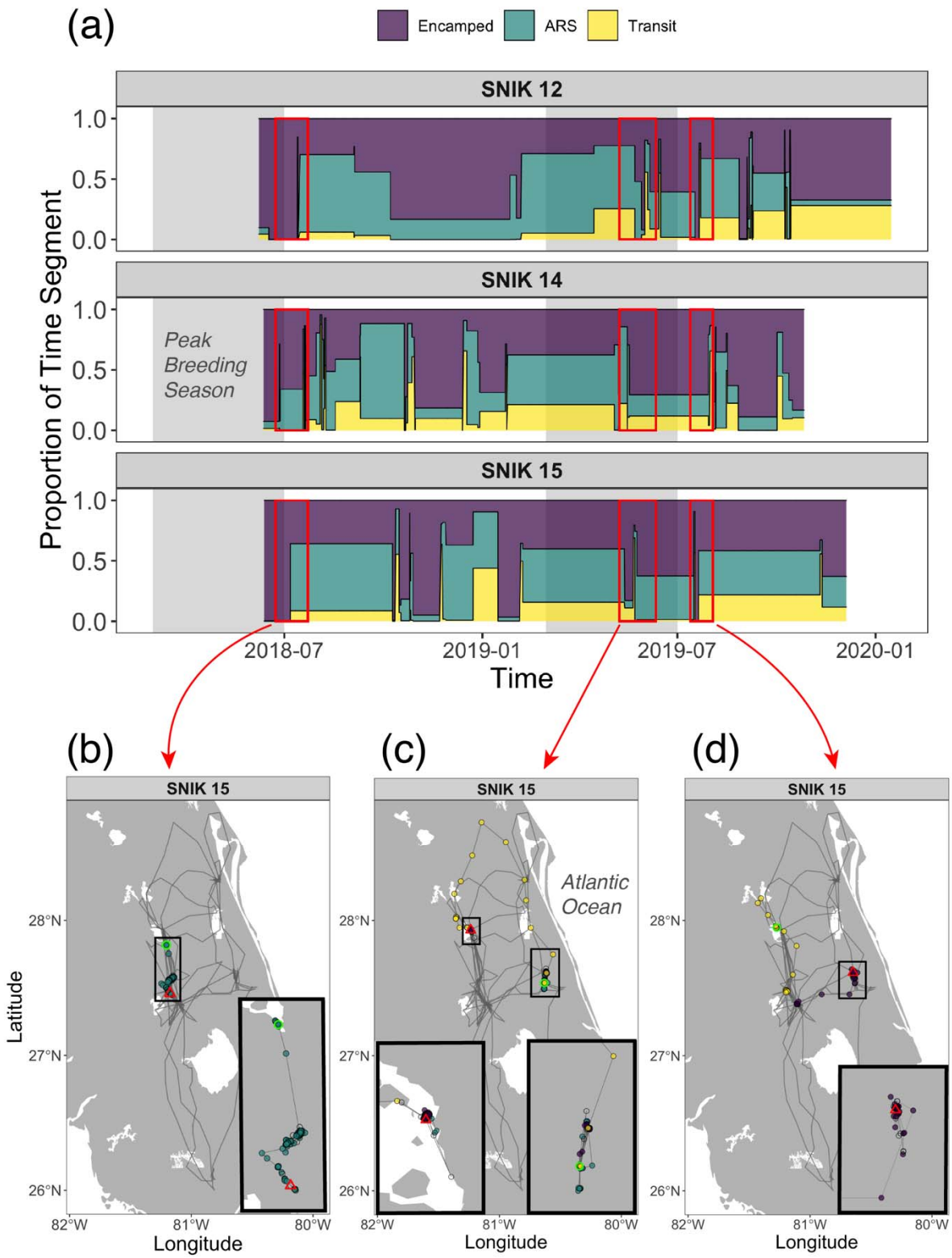


Figure 7 *Snail kite results from track segmentation and determination of the optimal number of behavioral states. (a) A subset of a time series from a single individual (SNIK 12) that displays estimated breakpoints (black lines) overlaying each data stream. Time series of step lengths and turning angles are shown on a continuous scale in relation to estimated breakpoints for ease of interpretation. (b) Boxplot showing the estimated proportions for each of the seven possible behaviors from all 439 track segments analyzed. (c) Distributions of step lengths and turning angles from each of the three retained behavioral states, ordered from slowest to fastest.*

Some individuals were only tracked for a short period of time and did not leave the natal area, however, 17 birds emigrated from their natal site. Dispersal events were typically denoted by a brief period of ARS or transit behavior (Figs 8a, b, S2). The three longest tracks, which belonged to snail kites tracked > 1 year (SNIK 12, SNIK 14, and SNIK 15), displayed relatively synchronous behavior before, during, and after their first breeding season. Two brief periods of high activity behavior that occurred during and immediately following peak breeding season in 2019 may potentially represent pre- and post-breeding dispersal events (Figs 8a, c, d).


(a)
Encamped
ARS
Transit
SNIK 12
SNIK 14
SNIK 15
Peak Breeding Season
Proportion of Time Segment
1.0
0.5
0.0
2018-07
2019-01
2019-07
2020-01
Time
(b)
(c)
(d)
Atlantic Ocean
Latitude
Longitude
28°N
27°N
26°N
82°W
81°W
80°W

444

Figure 8 *Patterns of movement behavior are shown over time for the three snail kites that were tagged over the longest durations, with particular focus on SNIK 15. (a) Time series of behavior proportions are shown with respect to peak breeding season (grey panels) for each of the three snail kites, with possible natal, pre-breeding, and post-breeding dispersal events denoted by red boxes. Maps of SNIK 15 movement depict the starting (green circle) and ending locations (red triangle) for (b) natal, (c) pre-breeding, and (d) post-breeding season dispersal events, as well as the dominant behavior associated with each track segment. Observations without behavior estimates (since not recorded at primary time interval) are shown as open points.*

4. Discussion

We demonstrated that our Bayesian framework (available within the `bayesmove` R package) can accurately identify changes in behavioral states, reliably estimate the optimal number of behavioral states, and properly characterize the state-dependent distributions of data streams. This two-stage model treats track segments as the unit of interest (as opposed to observations) and relies on the discretization of data streams to avoid the need to specify parametric probability distributions. Importantly, the proposed method is relatively computationally efficient, a key characteristic given the ever-increasing storage capacities of modern sensors and their ability to measure a growing number of intrinsic and environmental variables (Whitford and Klimley 2019; Williams et al. 2020). A comparison of model performance in addition to the analysis of an empirical dataset highlight the utility of our Bayesian framework.

4.1 Method Comparison

Although the BCPA was notably faster at track segmentation than the Bayesian method, accuracy of the estimated breakpoints was much higher with the latter. More specifically, BCPA estimated a greater number of breakpoints than there were true breakpoints for all but the longest

simulated tracks. Since BCPA has been shown to perform effectively in other studies (Gurarie et al. 2016; Nicholson et al. 2019), it is possible that the BCPA's parameters needed to be tuned individually for each of the simulated tracks as opposed to using a single set of values for all parameters. Alternatively, it is possible that BCPA's reliance on a single derived variable (persistence velocity) instead of separate step length and turning angle data streams explains the low performance that we found.

While HMMs are powerful methods that can incorporate many types of model structure (i.e., individual-level random effects, covariates, accounting for cyclical patterns) (Patterson et al. 2017; McClintock and Michelot 2018), they can also be restrictive in some of their assumptions. Standard forms of HMMs require the use of parametric distributions, which may not fit the data streams well (Langrock et al. 2018). In particular, the performance of HMMs can quickly deteriorate when the true underlying distributions are not well characterized by the selected parametric distributions. However, preliminary analyses demonstrated that HMMs display better performance than the Bayesian framework when the selected parametric distributions match the true underlying distributions (JA Cullen, unpublished data).

The determination of the optimal number of states is another issue when fitting HMMs since this is typically unknown *a priori* and is directly impacted by how well the selected parametric distributions characterize the states (Pohle et al. 2017). Unfortunately, HMMs often require multiple models to be fit and compared using information theoretic approaches, which tend to favor a greater number of states than are truly present and come at a high computational cost (Li and Bolker 2017; Pohle et al. 2017). These issues are directly addressed by our framework since we use a mixed-membership model (LDA) to cluster track segments, which performed much faster and more accurately than HMMs. Although hidden semi-Markov models

also rely on segment-level behavior, they are not easily implemented by available R packages and would require custom computer code (Joo et al. 2013; van de Kerk et al. 2015). While these existing methods can provide useful behavioral inference depending upon the ecological question and dataset, our framework provides a powerful alternative when behaviors are complex and data streams are not well characterized by parametric distributions.

4.2 Empirical applications

Three behavioral states were clearly estimated for the snail kite dataset, which was supported by biologically relevant distributions of step lengths and turning angles. The 'encamped' state likely represents fine-scale behaviors that include resting, feeding, and time spent at the nest (as a fledgling or reproductive adult). Since the distribution of step lengths characterizing the 'encamped' state are almost entirely < 30 m, which is the maximum location error for the tags, it is possible that this state represents observations where the individual is not moving at all. On the other hand, the 'ARS' state likely includes exploration for nearby suitable habitat as well as foraging bouts (Martin et al. 2006; Pias et al. 2016). Finally, the 'transit' state includes fast, directed movements associated with dispersal of snail kites in addition to departure from wetlands experiencing low water levels (Robertson et al. 2017).

The time series of snail kite behavior proportions showed variability in the timing of emigration from natal sites among individuals, but changes in behavior were generally synchronous in the three birds that reached maturity. This variability in the timing of emigration from natal sites could be due to a variety of factors, such as hatching date, body condition, and local environmental conditions (Rodgers and Schwikert 2003; Fletcher et al. 2015; Cattau et al. 2016). The shifts in behavior proportions appeared to show multiple phases of high and low

activity, some of which seem to match the phenology of natal dispersal (summer), pre-breeding dispersal (early-spring), and post-breeding dispersal (late-summer) (Bennetts and Kitchens 2000). While the continued monitoring of these tagged birds should provide greater evidence for the characterization of activity budgets over ontogeny, future research could also explore the primary drivers of snail kite movement and habitat use within each behavioral state through the inclusion of environmental covariates.

4.3 Caveats and extensions

Although the Bayesian framework effectively classified behavioral states from both simulated and empirical tracks, there are some limitations to this approach. The selection of the number and width of bins during the discretization of data streams is a subjective choice that impacts the results from the segmentation model and ultimately the estimation of behavioral states. Additionally, our model does not currently take location error into account, implicitly assuming that location error is negligible or requiring that it be accounted for via another method. Although our model can analyze data streams from regular or irregular time intervals, this will also depend on the inherent properties of the data streams themselves. Since step lengths and turning angles are calculated from multiple successive observations, these values will not be comparable once the data are not close to a regular time interval. However, variables such as net squared displacement (the squared distance from the starting location to all other relocations) can be analyzed over irregular time intervals.

This Bayesian framework can be extended to analyze other types of data streams and can include prior knowledge on the timing of behavioral shifts. Although only step lengths and turning angles were analyzed for the simulated and empirical tracks, additional ancillary data

coming from the sensor (e.g., elevation, salinity, temperature, or accelerometer data) can be used to make behavioral inference. These data streams can come from all types of distributions (i.e., continuous, discrete, bounded between 0 and 1). It is also relatively straightforward to deal with zero-inflated data by including all zeroes in a single bin. Additionally, our segmentation model can be implemented in a semi-supervised fashion, by which practitioners pre-specify breakpoints for the time series based on *a priori* knowledge and these breakpoints will be considered by the RJMCMC algorithm. This may be particularly useful if a single state can be clearly identified, but others may be more difficult to classify.

Acknowledgments

This work was supported in part by the University of Florida Biodiversity Institute (UFBI), the US Department of Agriculture's National Institute of Food and Agriculture (USDA-NIFA) McIntire–Stennis project 1005163, the US National Science Foundation (NSF) award 1458034, the Florida Fish and Wildlife Conservation Commission (FWC), and by the US Army Corp of Engineers (USACE). The tagging of snail kites was approved and conducted under UF IACUC no. 201708334. JAC was supported by a postdoctoral fellowship from the University of Florida Informatics Institute (UFII) Fellowship Program. We would like to thank Tyler Beck for his assistance with field work and Brenda Betancourt for her feedback during model and package development. Additionally, we are grateful to Rocío Joo and Mathieu Basille for their valuable comments on an earlier draft of this paper.

Authors' Contributions

J.A.C. and D.V. conceived the ideas and designed methodology; C.L.P. and R.J.F. collected the empirical data; J.A.C., C.L.P., and D.V. analyzed the data; J.A.C. and D.V. led the writing of the manuscript. All authors contributed critically to the drafts and gave final approval for publication.

Data Availability

All code mentioned here for the Bayesian framework is available within the bayesmove package for R hosted on CRAN at https://CRAN.R-project.org/package=bayesmove. The development version of the package is available on GitHub at https://github.com/joshcullen/bayesmove and a full set of vignettes can be found at https://joshcullen.github.io/bayesmove. The code to generate the simulations and perform method comparison are available on Zenodo (https://doi.org/10.5281/zenodo.4245255). The Florida snail kite telemetry data have not been made available since this is a federally listed endangered species and the location data are sensitive.

References

Bennetts RE, Kitchens WM. 2000. Factors influencing movement probabilities of a nomadic food specialist: proximate foraging benefits or ultimate gains from exploration? *Oikos* 91:459–467. doi:10.1034/j.1600-0706.2000.910306.x.

Blei DM, Ng AY, Jordan MI. 2003. Latent Dirichlet Allocation. *Journal of Machine Learning Research* 3:993–1022.

Burnham KP, Anderson DR. 2002. Model selection and multimodel inference: A practical information-theoretic approach. 2nd ed. New York: Springer.

Cattau CE, Fletcher RJ, Reichert BE, Kitchens WM. 2016. Counteracting effects of a non-native prey on the demography of a native predator culminate in positive population growth. *Ecological Applications* 26:1952–1968. doi:10.1890/15-1020.1/suppinfo.

Cullen JA, Valle D. 2020. bayesmove: Non-Parametric Bayesian Analysis of Animal Movement. R Package version 0.1.0.

Denison DGT, Holmes CC, Mallick BK, Smith AFM. 2002. Bayesian methods for nonlinear classification and regression. Chichester, UK: John Wiley & Sons.

Edelhoff H, Signer J, Balkenhol N. 2016. Path segmentation for beginners: an overview of current methods for detecting changes in animal movement patterns. *Movement Ecology* 4:21. doi:10.1186/s40462-016-0086-5.

Fletcher RJ, Robertson EP, Wilcox RC, Reichert BE, Austin JD, Kitchens WM. 2015. Affinity for natal environments by dispersers impacts reproduction and explains geographical structure of a highly mobile bird. *Proceedings of the Royal Society B: Biological Sciences* 282:20151545. doi:10.1098/rspb.2015.1545.

Fraser KC, Davies KTA, Davy CM, Ford AT, Flockhart DTT, Martins EG. 2018. Tracking the conservation promise of movement ecology. *Frontiers in Ecology and Evolution* 6:150. doi:10.3389/fevo.2018.00150.

Green PJ. 1995. Reversible jump Markov chain Monte Carlo computation and Bayesian model determination. *Biometrika* 82:711–732. doi:10.2307/2337340.

Gurarie E. 2014. bcpa: Behavioral change point analysis of animal movement. R package version 1.1.

Gurarie E, Andrews RD, Laidre KL. 2009. A novel method for identifying behavioural changes in animal movement data. *Ecology Letters* 12:395–408. doi:10.1111/j.1461-0248.2009.01293.x.

Gurarie E, Bracis C, Delgado M, Meckley TD, Kojola I, Wagner CM. 2016. What is the animal doing? Tools for exploring behavioural structure in animal movements. *Journal of Animal Ecology* 85:69–84. doi:10.1111/1365-2656.12379.

Hussey NE, Kessel ST, Aarestrup K, Cooke SJ, Cowley PD, Fisk AT, Harcourt RG, Holland KN, Iverson SJ, Kocik JF, et al. 2015. Aquatic animal telemetry: A panoramic window into the underwater world. *Science* 348:1255642. doi:10.1126/science.1255642.

Jonsen I. 2016. Joint estimation over multiple individuals improves behavioural state inference from animal movement data. *Scientific Reports* 6:20625. doi:10.1038/srep20625.

Jonsen ID, Basson M, Bestley S, Bravington MV, Patterson T a., Pedersen MW, Thomson R, Thygesen UH, Wotherspoon SJ. 2013. State-space models for bio-loggers: A methodological road map. *Deep Sea Research Part II: Topical Studies in Oceanography* 88–89:34–46. doi:10.1016/j.dsr2.2012.07.008. [accessed 2014 Jul 15]. http://linkinghub.elsevier.com/retrieve/pii/S096706451200094X.

Jonsen ID, Mills Flemming J, Myers RA. 2005. Robust state-space modeling of animal movement data. *Ecology* 86:2874–2880. doi:10.1890/04-1852.

Joo R, Bertrand S, Tam J, Fablet R. 2013. Hidden Markov models: The best models for forager movements? *PLoS ONE* 8:e71246. doi:10.1371/journal.pone.0071246.

Joo R, Picardi S, Boone ME, Clay TA, Patrick SC, Romero-Romero VS, Basille M. 2020. A decade of movement ecology. *arXiv preprint*:2006.00110.

Kays R, Crofoot MC, Jetz W, Wikelski M. 2015. Terrestrial animal tracking as an eye on life and planet. *Science* 348:aaa2478. doi:10.1126/science.aaa2478.

van de Kerk M, Onorato DP, Criffield MA, Bolker BM, Augustine BC, Mckinley SA, Oli MK. 2015. Hidden semi-Markov models reveal multiphasic movement of the endangered Florida

panther. *Journal of Animal Ecology* 84:576–585. doi:10.1111/1365-2656.12290.

Langrock R, Adam T, Leos-Barajas V, Mews S, Miller DL, Papastamatiou YP. 2018. Spline-based nonparametric inference in general state-switching models. *Statistica Neerlandica* 72:179–200. doi:10.1111/stan.12133.

Langrock R, King R, Matthiopoulos J, Thomas L, Fortin D, Morales JM. 2012. Flexible and practical modeling of animal telemetry data: hidden Markov models and extensions. *Ecology* 93:2336–2342. doi:10.1890/11-2241.1.

Li M, Bolker BM. 2017. Incorporating periodic variability in hidden Markov models for animal movement. *Movement Ecology* 5:1. doi:10.1186/s40462-016-0093-6.

Martin J, Nichols JD, Kitchens WM, Hines JE. 2006. Multiscale patterns of movement in fragmented landscapes and consequences on demography of the snail kite in Florida. :527–539. doi:10.1111/j.1365-2656.2006.01073.x.

McClintock BT, Michelot T. 2018. momentuHMM: R package for generalized hidden Markov models of animal movement. *Methods in Ecology and Evolution* 9:1518–1530. doi:10.1111/2041-210X.12995.

Michelot T, Langrock R, Patterson TA. 2016. moveHMM: an R package for the statistical modelling of animal movement data using hidden Markov models. *Methods in Ecology and Evolution* 7:1308–1315. doi:10.1111/2041-210X.12578.

Nathan R, Getz WM, Holyoak M, Kadmon R, Saltz D, Smouse PE. 2008. A movement ecology paradigm for unifying organismal movement research. *Proceedings of the National Academy of Sciences* 105:19052–19059. doi:10.1021/i360006a005.

Nicholson KL, Warren MJ, Rostan C, Månsson J, Paragi TF, Sand H. 2019. Using fine-scale movement patterns to infer ungulate parturition. *Ecological Indicators* 101:22–30.

doi:10.1016/j.ecolind.2019.01.004.

Patin R, Etienne MP, Lebarbier E, Chamaillé-Jammes S, Benhamou S. 2020. Identifying stationary phases in multivariate time series for highlighting behavioural modes and home range settlements. *Journal of Animal Ecology* 89:44–56. doi:10.1111/1365-2656.13105.

Patterson TA, Parton A, Langrock R, Blackwell PG, Thomas L, King R. 2017. Statistical modelling of individual animal movement: an overview of key methods and a discussion of practical challenges. *AStA Advances in Statistical Analysis* 101:399–438. doi:10.1007/s10182-017-0302-7.

Pias KE, Fletcher RJ, Kitchens WM. 2016. Assessing the value of novel habitats to snail kites through foraging behavior and nest survival. *Journal of Fish and Wildlife Management* 7:449–460. doi:10.3996/022016-JFWM-008.

Pohle J, Langrock R, van Beest FM, Schmidt NM. 2017. Selecting the number of states in hidden Markov models: Pragmatic solutions illustrated using animal movement. *Journal of Agricultural, Biological, and Environmental Statistics* 22:270–293. doi:10.1007/s13253-017-0283-8.

Potts JR, Börger L, Scantlebury DM, Bennett NC, Alagaili A, Wilson RP. 2018. Finding turning-points in ultra-high-resolution animal movement data. *Methods in Ecology and Evolution* 9:2091–2101. doi:10.1111/2041-210X.13056.

Reichert B, Cattau C, Fletcher Jr. R, Sykes Jr. P, Rodgers Jr. J, Bennetts R. 2020. Snail Kite (Rostrhamus sociabilis), version 1.0. In: Poole A, editor. Birds of the World. Ithaca, NY, USA: Cornell Lab of Ornithology.

Robertson EP, Fletcher RJ, Austin JD. 2017. The causes of dispersal and the cost of carry-over effects for an endangered bird in a dynamic wetland landscape. *Journal of Animal Ecology*

86:857–865. doi:10.1111/1365-2656.12676.

Rodgers JA, Schwikert ST. 2003. Breeding chronology of snail kites (Rostrhamus sociabilis plumbeus) in central and south Florida wetlands. *Southeastern Naturalist* 2:293–300. doi:10.1656/1528-7092(2003)002[0293:bcoskr]2.0.co;2.

Seidel DP, Dougherty E, Carlson C, Getz WM. 2018. Ecological metrics and methods for GPS movement data. *International Journal of Geographical Information Science* 32:2272–2293. doi:10.1080/13658816.2018.1498097.

Sisson SA. 2005. Transdimensional Markov chains: A decade of progress and future perspectives. *Journal of the American Statistical Association* 100:1077–1089. doi:10.1198/016214505000000664.

Valle D, Baiser B, Woodall CW, Chazdon R. 2014. Decomposing biodiversity data using the Latent Dirichlet Allocation model, a probabilistic multivariate statistical method. *Ecology Letters* 17:1591–1601. doi:10.1111/ele.12380.

Valle D, Cvetojevic S, Robertson EP, Reichert BE, Hochmair HH, Fletcher RJ. 2017. Individual movement strategies revealed through novel clustering of emergent movement patterns. *Scientific Reports* 7:44052. doi:10.1038/srep44052.

Whitford M, Klimley AP. 2019. An overview of behavioral, physiological, and environmental sensors used in animal biotelemetry and biologging studies. *Animal Biotelemetry* 7:26. doi:10.1186/s40317-019-0189-z.

Williams HJ, Taylor LA, Benhamou S, Bijleveld AI, Clay TA, de Grissac S, Demšar U, English HM, Franconi N, Gómez-Laich A, et al. 2020. Optimizing the use of biologgers for movement ecology research. *Journal of Animal Ecology* 89:186–206. doi:10.1111/1365-2656.13094.

Wilmers CC, Nickel B, Bryce CM, Smith JA, Wheat RE, Yovovich V. 2015. The golden age of

bio-logging: how animal-borne sensors are advancing the frontiers of ecology. *Ecology* 96:1741–1753. doi:https://doi.org/10.1890/14-1401.1.