Incorporating Telemetry Error into Hidden Markov Models of Animal Movement Using Multiple Imputation

Brett T. MCCLINTOCK

When data streams are observed without error and at regular time intervals, discrete-time hidden Markov models (HMMs) have become immensely popular for the analysis of animal location and auxiliary biotelemetry data. However, measurement error and temporally irregular data are often pervasive in telemetry studies, particularly in marine systems. While relatively small amounts of missing data that are missing-completely-at-random are not typically problematic in HMMs, temporal irregularity can result in few (if any) observations aligning with the regular time steps required by HMMs. Fitting HMMs that explicitly account for uncertainty attributable to location measurement error, temporally irregular observations, or other forms of missing data typically requires computationally demanding techniques, such as Markov chain Monte Carlo (MCMC). Using simulation and a real-world bearded seal (*Erignathus barbatus*) example, I investigate a practical alternative to incorporating measurement error and temporally irregular observations into HMMs based on multiple imputation of the position process drawn from a single-state continuous-time movement model. This two-stage approach is relatively simple, performed with existing software using efficient maximum likelihood methods, and completely parallelizable. I generally found the approach to perform well across a broad range of simulated measurement error and irregular sampling rates, with latent states and locations reliably recovered in nearly all simulated scenarios. However, high measurement error coupled with low sampling rates often induced bias in both the estimated probability distributions of data streams derived from the imputed position process and the estimated effects of spatial covariates on state transition probabilities. Results from the two-stage analysis of the bearded seal data were similar to a more computationally intensive single-stage MCMC analysis, but the two-stage analysis required much less computation time and no custom model-fitting algorithms. I thus found the two-stage multiple-imputation approach to be promising in terms of its ease of implementation, computation time, and performance. Code for implementing the approach using the R package "momentuHMM" is provided.
Supplementary materials accompanying this paper appear online.

Key Words: Argos satellite telemetry; Bearded seal; Biotelemetry; crawl; momentuHMM; moveHMM; Multistate models.

Brett T. McClintock (✉) Marine Mammal Laboratory, Alaska Fisheries Science Center, NOAA-NMFS, 7600 Sand Point Way NE, Seattle, WA 98115, USA (E-mail: *brett.mcclintock@noaa.gov*).

1. INTRODUCTION

Discrete-time hidden Markov models (HMMs) have become immensely popular for the analysis of animal telemetry data (e.g., Morales et al. 2004; Jonsen et al. 2005; Langrock et al. 2012; McClintock et al. 2012). In short, an HMM is a time series model composed of one or more observable data streams $(X_1, \ldots, X_T)$, each of which is generated by Z state-dependent probability distributions, where the unobservable (hidden) state sequence ($z_t \in \{1, \ldots, Z\}$ for $t = 1, \ldots, T$) is assumed to be a Markov chain. The state sequence of the Markov chain is governed by (typically first-order) state transition probabilities, $\gamma_{ij}^{(t)} = \Pr(z_{t+1} = j \mid z_t = i)$ for $i, j = 1, \ldots, Z$, and an initial distribution $\boldsymbol{\delta}^{(1)}$. The likelihood of an HMM can be succinctly expressed using the forward algorithm:

$$\mathcal{L} = \boldsymbol{\delta}^{(1)} \boldsymbol{\Gamma}^{(1)} \mathbf{P}(x_1) \, \boldsymbol{\Gamma}^{(2)} \mathbf{P}(x_2) \, \boldsymbol{\Gamma}^{(3)} \cdots \boldsymbol{\Gamma}^{(T-1)} \mathbf{P}(x_{T-1}) \, \boldsymbol{\Gamma}^{(T)} \mathbf{P}(x_T) \, \mathbf{1}^Z \qquad (1)$$

where the $Z \times Z$ transition probability matrix $\boldsymbol{\Gamma}^{(t)} = (\gamma_{ij}^{(t)})$, $\mathbf{P}(x_t) = \text{diag}(p_1(x_t), \ldots, p_Z(x_t))$, $p_s(x_t)$ is the conditional probability density of X_t given $z_t = s$, and $\mathbf{1}^Z$ is a Z-vector of ones (see Zucchini et al. 2016 for a thorough introduction to HMMs).

While HMMs for animal movement based solely on location data are somewhat limited in the number and type of biologically meaningful movement behavior states they are able to accurately identify (Morales et al. 2004; Beyer et al. 2013; Bagniewska et al. 2013; McClintock et al. 2014), multivariate HMMs that utilize both location and auxiliary biotelemetry data (e.g., McClintock et al. 2013; Russell et al. 2015; DeRuiter et al. 2016) can facilitate the identification of additional states that go beyond the two-state approaches that are most frequently used by practitioners ("encamped" and "exploratory" states *sensu* Morales et al. 2004 or "foraging" and "transit" states *sensu* Jonsen et al. 2005).

When data streams are observed without error and at regular time intervals, a major advantage of HMMs is the relatively fast and efficient maximization of the likelihood using the forward algorithm (Zucchini et al. 2016), and user-friendly software is available for fitting movement HMMs under these circumstances (e.g., Michelot et al. 2016a). However, location measurement error is rarely nonexistent and depends on both the telemetry device and the system under study. For example, GPS errors are typically less than 50m, but Argos errors can exceed 10 km (Costa et al. 2010; Silva et al. 2014). Missing location or auxiliary biotelemetry data typically arise when transmitters cannot communicate with satellites at temporally regular intervals. While missing data can be a problem in terrestrial systems (e.g., in canyons or dense forest), it can often be pervasive in marine environments. If some X_t are missing, these data gaps are often ignored in maximum likelihood analyses (by replacing $\mathbf{P}(x_t)$ with the $Z \times Z$ unity matrix) and thus do not contribute information to the estimation of state-dependent probability distribution parameters (e.g., Zucchini et al. 2016), but if data are frequently missing or not missing-completely-at-random, this strategy could have undesirable inferential consequences (Nakagawa and Freckleton 2008). Another strategy is to impute missing data using simple linear interpolation (e.g., Russell et al. 2015; DeRuiter et al. 2016; Michelot et al. 2016b), although the reliability of this approach is poorly understood. While relatively small amounts of missing data that are missing-completely-at-random are not typically problematic in HMMs, an extreme case of missing data can arise

when location data are obtained with little or no temporal regularity, as in many marine mammal telemetry studies (e.g., Jonsen et al. 2005), such that few (if any) observations align with the regular time steps required by discrete-time HMMs. Temporal irregularity is therefore different from the conventional HMM missing data scenario where data streams are consistently observed at regular time intervals, but some of these observations happen to be missing.

When explicitly accounting for uncertainty attributable to location measurement error, temporally irregular observations, or other forms of missing data, one must typically fit HMMs using computationally intensive (and often time-consuming) model-fitting techniques such as Markov chain Monte Carlo (Jonsen et al. 2005; McClintock et al. 2013; Turek et al. 2016). For example, a recent multivariate HMM analysis of bearded seal (*Erignathus barbatus*) telemetry data that incorporated six behavior states, seven data streams, location measurement error, temporal irregularity, and missing auxiliary data required several *weeks* to fit using MCMC (McClintock et al. 2017). Given the time and money typically expended in deploying animal-borne telemetry devices, one could posit that such an "expensive" analysis is entirely justified. However, complex analyses requiring novel statistical methods and custom model-fitting algorithms are not practical for many of the biologists and ecologists conducting these studies.

Here I investigate a practical approach to incorporating measurement error and temporal irregularity into HMMs for animal movement using multiple imputation (Rubin 1987; Hooten et al. 2017). This two-stage approach is relatively simple and can be performed with existing software using efficient maximum likelihood methods. After describing the approach in more detail in the next section, I investigate its performance properties in a series of simulation experiments. I then approximate the bearded seal analysis of McClintock et al. (2017) using multiple imputation and compare the results.

2. METHODS

The multiple-imputation approach for incorporating location measurement error and temporally irregular or missing observations into animal movement HMMs consists of two stages. The basic concept is to first employ a single-state (i.e., $Z = 1$) movement model that is relatively easy to fit but can accommodate location measurement error and temporally irregular or missing observations. The second stage involves repeatedly fitting an HMM to n temporally regular realizations of the position process drawn from the model output of the first stage. These temporally regular realizations of the position process constitute the data "we wish we had" if the observation process were not subject to location measurement error and temporally irregular or missing observations. Inferences about behavior states (e.g., state decoding, state probabilities, "activity budgets"), transition probabilities, and state-dependent probability distribution parameters are based on a pooling of the n imputed data HMM analyses using standard multiple-imputation formulae (Rubin 1987).

For the first stage of the analysis, the continuous-time correlated random walk model of Johnson et al. (2008) is very well suited and easily implemented in the R package "crawl" (Johnson 2016) using maximum likelihood methods, although any movement model that

accommodates both location measurement error and temporally irregular or missing observations could be fitted to the telemetry locations during this stage. Some advantages of the Johnson et al. (2008) model in "crawl" are its speed, flexible measurement error model specification, and ease of repeatedly drawing realizations of the temporally regular position process from the model output.

The second stage proceeds by fitting an HMM to each of the n imputed data sets using standard methods. The HMM data streams need not be limited to step length or turning angle (e.g., Morales et al. 2004), and auxiliary biotelemetry data streams that inform behavior states can also be included at this stage (e.g., dive activity, altitude, heart rate; see McClintock et al. 2013). Data streams or covariates that are dependent on location (e.g., step length, turning angle, habitat type, snow depth, sea surface temperature) will of course vary among the n realizations of the position process, and the pooled inferences across the HMM analyses will therefore reflect location uncertainty. Pooled point and variance estimates can be calculated as

$$\bar{\theta} = \frac{1}{n} \sum_{i=1}^{n} \theta^{(i)} \tag{2}$$

and

$$\text{var}\left(\bar{\theta}\right) = \left[\frac{1}{n} \sum_{i=1}^{n} \text{var}\left(\theta^{(i)}\right)\right] + \left(1 + \frac{1}{n}\right) \left[\frac{1}{n-1} \sum_{i=1}^{n} \left(\theta^{(i)} - \bar{\theta}\right)^2\right], \tag{3}$$

respectively, where $\theta^{(i)}$ and $\text{var}(\theta^{(i)})$ are the ith point and variance estimates for parameter θ (Rubin and Schenker 1986).

3. SIMULATION STUDY

3.1. SIMULATION METHODS

I performed three sets of simulation experiments to evaluate the performance of the multiple-imputation approach under a variety of ecological and sampling scenarios (Table 1). All simulated datasets consisted of $N = 7$ individual tracks generated from identical probability distributions for each data stream. The length of each track was between 500 and 1500 regularly spaced time steps, and the same "true" tracks were used for all scenarios within each set of simulations. The observed location data ($\mathbf{y}$) were generated from the "true" locations ($\boldsymbol{\mu}$) subject to varying levels of measurement error and temporal irregularity, but the underlying movement process model in each case was a standard HMM consisting of T regular time steps. Temporal irregularity was introduced by allowing observations ($\mathbf{y}_k$) to fall somewhere along the straight line between the temporally regular locations for each time step:

$$\mathbf{y}_k = (1 - j_k) \, \boldsymbol{\mu}_{t-1} + j_k \boldsymbol{\mu}_t + \boldsymbol{\epsilon}_k$$

for all $k \in (t-1, t]$ and $t = 2, \ldots, T + 1$, where $j_k \in (0, 1]$ is the proportion of the regular time interval between locations $\boldsymbol{\mu}_{t-1}$ and $\boldsymbol{\mu}_t$ at which y_k was observed, and $\boldsymbol{\epsilon}_k$ is (bivariate normal) location measurement error. The objective of the simulations was to assess whether or not the parameters, state sequences, and temporally regular locations of

Table 1.   Simulation inputs for evaluating performance of multiple imputation as an approach to incorporating location measurement error and temporally irregular observations into hidden Markov models of animal movement. Behavior states included "foraging" (F), "transit" (T), "hauled out on ice" (I), and "resting at sea" (S). State-dependent data streams included step length (s_z), turning angle (ϕ_z), number of dives (d_z), proportion dive time (w_z), and proportion dry time (v_z) for $z \in \{F, T, I, S\}$ and were generated from Gamma(shape, scale), wrapped Cauchy [wCauchy(mean, concentration)], Poisson(rate), or beta (shape1, shape2) distributions. Each simulation was based on $N = 7$ simulated individual tracks. Three sets of simulations were conducted. The "two-state with covariate" simulations were similar to the "two-state" simulations, but included a spatially correlated covariate on transition probabilities.

Data stream	Simulation set		
	Two-state	Two-state with covariate	Four-state
Step length	$s_F \sim$ Gamma(0.25, 400) $s_T \sim$ Gamma(25, 40)	$s_F \sim$ Gamma(0.25, 400) $s_T \sim$ Gamma(25, 40)	$s_F \sim$ Gamma(100, 5) $s_T \sim$ Gamma(100, 50) $s_I, s_S \sim$ Gamma(1, 10)
Turning angle	$\phi_F \sim$ wCauchy(0, 0.00) $\phi_T \sim$ wCauchy(0, 0.75)	$\phi_F \sim$ wCauchy(0, 0.00) $\phi_T \sim$ wCauchy(0, 0.75)	$\phi_F \sim$ wCauchy(0, 0.10) $\phi_T \sim$ wCauchy(0, 0.75) $\phi_I, \phi_S \sim$ wCauchy(0, 0.85)
No. of dives			$d_F \sim$ Poisson(20) $d_T \sim$ Poisson(5) $d_I, d_S \sim$ Poisson(1)
Dive time			$w_F, w_T \sim$ Beta(10, 1) $w_I, w_S \sim$ Beta(1, 10)
Dry time			$v_F, v_T, v_S \sim$ Beta(1, 10) $v_I \sim$ Beta(10, 1)

the "true" HMM could be reliably recovered from the observed data ($\mathbf{y}$) using the proposed two-stage approach.

The first set used simulated location data typical of the most commonly employed two-state HMMs of animal movement (e.g., Morales et al. 2004; Jonsen et al. 2005) consisting of an area-restricted-search-type state (i.e., "encamped" or "foraging") and a high-speed, directionally persistent state (i.e., "exploratory" or "transit"). Based on the results of Beyer et al. (2013) demonstrating HMMs can perform poorly when movement behavior states are not sufficiently distinct, the probability distributions for step length and turning angle were chosen such that they had little overlap between states (Table 1). The transition probability matrix for these simulations was $\mathbf{\Gamma} = \begin{bmatrix} 0.8 & 0.2 \\ 0.1 & 0.9 \end{bmatrix}$, where each element $\gamma_{i,j}$ is the probability of switching from state i at time t to state j at time $t + 1$ for $i, j \in \{1 = \text{"foraging"}, 2 = \text{"transit"}\}$. Temporal irregularity was simulated by assuming the wait times between the observed locations followed an exponential distribution with rate λ, which can be interpreted as the expected number of (temporally irregular) locations observed between (temporally regular) time steps t and $t + 1$. I limited the design points to $\lambda = 2, 1$, and 0.5, but also included a design point assuming temporal regularity (i.e., $T + 1$ observations occurring at temporally regular times $t = 1, \ldots, T + 1$). Location measurement error was assumed to arise from a bivariate normal distribution based on an error ellipse with semi-major axis of length M, semi-minor axis of length m, and orientation c (McClintock et al. 2015). For simplicity, I assumed $M = m, c = 0$, and limited the design

points to $M = 50$ (low measurement error relative to state-dependent scales of movement), $M = 500$ (moderate error), $M = 1500$ (high error), and $M = 3000$ (extreme error).

The second set of simulations were identical to the first except for one key difference. For these scenarios, state transition probabilities were assumed to be a function of a spatially correlated covariate, whereby simulated individuals were more likely to switch to (and remain in) the "foraging" state ($z_t = 1$) at locations with larger covariate values. The transition probability matrix for these simulations was $\mathbf{\Gamma}^{(t)} = \begin{bmatrix} 1 - \gamma_{1,2}^{(t)} & \gamma_{1,2}^{(t)} \\ \gamma_{2,1}^{(t)} & 1 - \gamma_{2,1}^{(t)} \end{bmatrix}$, where $\gamma_{1,2}^{(t)} = \text{logit}^{-1}(5 - 25c_t)$, $\gamma_{2,1}^{(t)} = \text{logit}^{-1}(-10 + 50c_t)$, and $c_t \in [0, 1]$ is the spatial covariate value corresponding to the animal's location at time t.

The third set of simulations was motivated by multivariate HMMs that utilize both location and additional data streams to inform behavioral states of ice-associated seals (e.g., McClintock et al. 2017). Data were generated under $Z = 4$ behavior states (1 = "hauled out on ice," 2 = "resting at sea," 3 = "foraging," and 4 = "transit") and characterized by 6 data streams: step length, turning angle, number of dives, proportion dive time, proportion dry time, and proportion sea ice cover (see Table 1). Because the step length distribution, turning angle distribution, and transition probabilities for the "hauled out on ice" and "resting at sea" states were identical, I simply simulated state sequences for three states $\left(\mathbf{\Gamma} = \begin{bmatrix} 0.65 & 0.35 & 0.01 \\ 0.09 & 0.82 & 0.09 \\ 0.04 & 0.24 & 0.72 \end{bmatrix} \right)$ and then delineated these two non-diving states based on whether or not the sea ice concentration at the initial location of the corresponding time step was > 0.05. A spatially correlated sea ice concentration grid was simulated using the R package "gstat" (Pebesma 2004). The second stage of the multiple-imputation approach requires more computation time with four states, so I limited the measurement error scenarios to $M \in \{50, 500, 1500\}$ for this set of simulations.

The same procedure was followed for all scenarios within the three sets of simulations:

(1) Separately fit the continuous-time correlated random walk movement model of Johnson et al. (2008) to the observed data ($\mathbf{y}$) for each of the $N = 7$ individuals using the crwMLE() function in the R package "crawl" (Johnson 2016). Assume a bivariate normal error ellipse model for each observed location $\mathbf{y}_k \sim N(\mathbf{\mu}_k, \mathbf{\Sigma}_k)$, where $\mathbf{\mu}_k$ is the true location and $\mathbf{\Sigma}_k = \begin{pmatrix} \frac{M^2}{2} & 0 \\ 0 & \frac{M^2}{2} \end{pmatrix}$ at time $k \in [1, T + 1]$.

(2) Use the "crawl" function crwPostIS() to draw n samples from the posterior distribution of the position process $\left(\mathbf{\mu}_t^{(i)}, i = 1, \ldots, n \right)$ at temporally regular intervals ($t = 1, 2, \ldots, T + 1$) conditional on the fitted parameters for each individual from step 1. For each of the n realizations of the position process, calculate step lengths, turning angles, and any other data streams or covariates that depend on location.

(3) Estimate the model parameters by fitting the corresponding HMM (consisting of the "Data stream" distributions in Table 1) to each of the n imputed data sets using maximum likelihood methods. For each of the n HMM fits, use the Viterbi algorithm

to estimate the most likely state sequence and the forward–backward algorithm to estimate state probabilities for each time step (Zucchini et al. 2016).

(4) Use Eqs. (2) and (3) to calculate pooled parameter estimates, variances, and 95% t-distributed confidence intervals (Rubin and Schenker 1986).

Except for the two-state model including a spatial covariate on state transition probabilities, the initial distribution was assumed to be the stationary distribution for the fitted HMMs. Note that the Viterbi and forward–backward algorithms do not provide estimates of uncertainty. For the "pooled" estimates of the most likely state sequence ($z_t, t = 1, \ldots, T$), I simply calculated the mode for each time step from the output of the Viterbi algorithm for each of the n HMM fits. For "pooled" point and variance estimates of the state probabilities for each time step ($q_{z,t}, z \in \{1, \ldots, Z\}, t = 1, \ldots, T$), I respectively used Eqs. (2) and (3) assuming $1/n \sum_{i=1}^{n} \text{var}(\theta^{(i)}) = 0$.

When fitting the four-state HMMs, I included the imputed sea ice concentration data as an additional data stream modeled as a (state-dependent) beta distribution. This of course is not how the data were generated, and a more sophisticated approach would be to incorporate sea ice concentration into the transition probabilities for the non-diving states ("hauled out on ice" and "resting at sea"), possibly using a hidden semi-Markov model (Zucchini et al. 2016). However, hidden semi-Markov models are more challenging to implement than HMMs, and my goal was to evaluate this approximation as a similar strategy was used by McClintock et al. (2017) to help distinguish different types of resting behavior in bearded seals.

I evaluated the performance of the multiple-imputation approach based on its ability to estimate the unobserved state sequence, the parameters of the state-dependent probability distributions, and the temporally regular locations ($\boldsymbol{\mu}_t$). Classification accuracy was evaluated based on the proportion of time steps for the estimated most likely state sequence that matched the true state, a measure of agreement between the true and estimated states that takes chance agreement into account (Congalton 1991), and the proportion of estimated state probabilities ($q_{z,t}$) with at least 0.05 probability assigned to the true state. I also compared the multiple-imputation approach to an HMM fitted to the single most likely position process predicted by each of the "crawl" model fits (hereafter referred to as "the single-imputation approach"). For all three sets of simulations, I first fit the single-imputation model using the true parameters as starting values and then used the estimated parameters from the single-imputation model fit as starting values for the multiple-imputation model fits. I used similar parameter constraints to those used by McClintock et al. (2017) to avoid label switching among the n HMMs fitted for each simulation scenario.

For all scenarios within each set of simulations, $n = 400$ imputations were performed. To provide some insight into the number of imputations that may be required in practice, pooled estimates from the $n = 400$ model fits were compared to those of randomly selected subsets of $n = 30$ and $n = 5$ imputations. All analyses were performed in R (R Core Team 2016) using an extension of the "moveHMM" package (Michelot et al. 2016a) that is currently under development (https://github.com/bmcclintock/momentuHMM). The "momentuHMM" package allows for additional data streams, as well as user-specified

design matrices and constraints for the state-dependent probability distribution parameters. R code for simulating data and implementing the two-stage analyses using "momentuHMM" is provided in ESM of Appendix A.

3.2. SIMULATION RESULTS

For the two-state simulations without a spatial covariate on transition probabilities ($T = 6519$), each individual HMM typically required about 13 s to converge. All $n = 400$ imputed data analyses required about 23 min (range $= 16 - 27$ min) when run in parallel on a 3.7 GHz processor with eight cores. I generally found the multiple-imputation approach to perform well in terms of state and probability distribution estimation. The single-imputation approach performed well with high sampling rates and low measurement error, but was less reliable with lower sampling rates or higher measurement error (Fig. 1; ESM of Appendix B). The estimated most likely state sequence ($z_t, t = 1, \ldots, T$) was typically similar between the single- and multiple-imputation approaches, and these became less accurate as measurement error increased or sampling rate decreased. However, unlike the single-imputation approach, the estimated state probabilities ($q_{z,t}$) for the multiple-imputation approach almost always included at least a 5% probability for the true state at each time step (Fig. 1). By committing few such "false-positive" state assignments, the multiple-imputation approach proved much more reliable in characterizing state assignment uncertainty attributable to high measurement error or low sampling rates. Both the single- and multiple-imputation approaches were increasingly unable to recover the true step length (Fig. 1) and turning angle (ESM of Appendix B) distributions as measurement error increased and, to a somewhat lesser extent, as sampling rate decreased. In terms of path reconstruction, the multiple-imputation approach tended to estimate 95% confidence bands for $\boldsymbol{\mu}_t$ that included the true value (Fig. 1). However, mean coverage of $\boldsymbol{\mu}$ when $M = 50$ m was as low as 80, 85, and 88% for $\lambda = 2, 1$, and 0.5 (Fig. 1), and the mean errors in the estimated locations were 88.0, 164.7, and 345.5 m, respectively.

For the two-state simulations that included a spatial covariate on state transition probabilities ($T = 7000$), each individual HMM typically required about 20 s (range $= 6$–40 s) to converge. All $n = 400$ imputed data analyses required about 19 min (range $= 12$–24 min) when run in parallel on a 3.7 GHz processor with seven cores. Both the single- and multiple-imputation approaches generally performed well, and in most cases the inclusion of a spatial covariate on state transition probabilities made the models more robust to location measurement error and temporal irregularity relative to the two-state scenarios without a covariate. Across all scenarios, the multiple-imputation approach tended to outperform the single-imputation approach in terms of state (Fig. 2) and probability density estimation (ESM of Appendix B). As with the two-state simulations without a covariate, both approaches were increasingly unable to recover the true step length and turning angle distributions as measurement error increased and sampling rate decreased. Similarly, the effect of the spatial covariate on transition probabilities was increasingly underestimated as measurement error increased and sampling rate decreased (ESM of Appendix B). The multiple-imputation approach tended to estimate 95% confidence bands that covered $\boldsymbol{\mu}_t$, but mean coverage when $M = 50$ m was as low as 81, 84, and 86% for $\lambda = 2, 1$, and 0.5
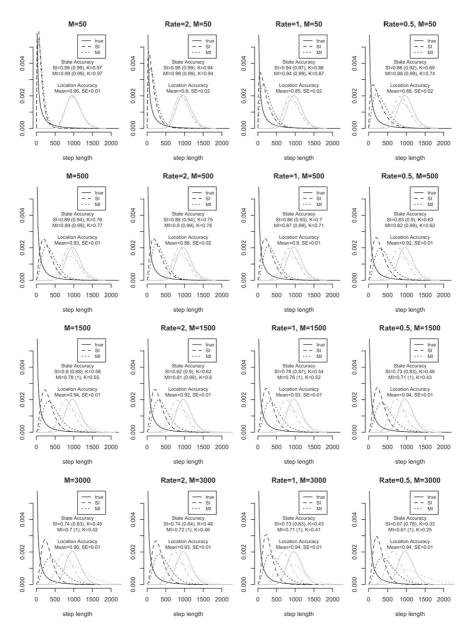
Figure 1. Selected simulation results under the two-state ("foraging" in *black* and "transit" in *gray*) scenarios with no spatial covariate on state transition probabilities. Each panel presents the true and estimated probability densities for step length based on the single-imputation (SI) and multiple-imputation (MI) approaches with varying levels of location measurement error ($M \in \{50, 500, 1500, 3000\}$) and temporally irregular sampling (Rate $\in$ 2, 1, 0.5). The first column pertains to scenarios with temporally regular sampling that exactly matches the time steps of the simulated tracks. State classification accuracy for each scenario is based on the proportion of time steps for the estimated most likely state sequence that match the true state, the proportion of estimated state probabilities with at least 0.05 probability assigned to the true state (in parentheses), and a measure of agreement between the true and estimated states (K) that takes chance agreement into account and ranges from 0 (no better than chance) to 1 (perfect agreement not attributable to chance). Location accuracy for each scenario is based on the proportion of true locations that fall within their respective estimated 95% confidence bands using the multiple-imputation approach. Multiple-imputation pooled estimates are based on $n = 400$ realizations of the position process.
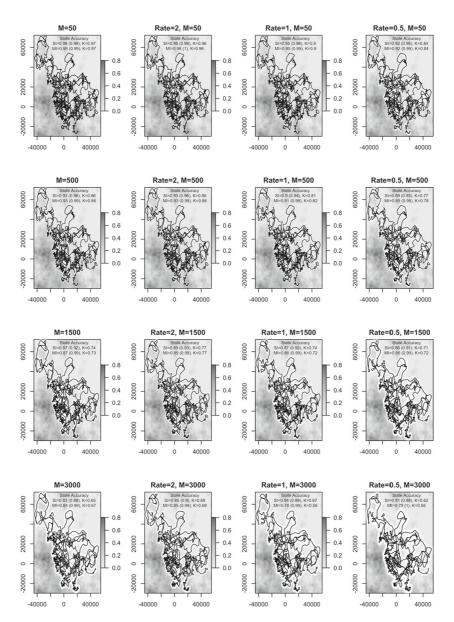
Figure 2.   Selected simulation results under the two-state ("foraging" and "transit") scenarios including a spatial covariate on state transition probabilities. Each panel presents the estimated tracks for $N = 7$ individuals relative to the spatially correlated covariate under varying levels of location measurement error ($M \in \{50, 500, 1500, 3000\}$) and temporally irregular sampling (Rate $\in 2, 1, 0.5$). The first column pertains to scenarios with temporally regular sampling that exactly matches the time steps of the simulated tracks. *Black lines* indicate the tracks using the single-imputation (SI) approach, while *white lines* indicate each realization of the position process using the multiple-imputation (MI) approach. Under the data-generating model, higher covariate values were associated with higher probabilities of switching to (and remaining in) the "foraging" state characterized by area-restricted-search-type movement. State classification accuracy for each scenario is based on the proportion of time steps for the estimated most likely state sequence that match the true state, the proportion of estimated state probabilities with at least 0.05 probability assigned to the true state (in parentheses), and a measure of agreement between the true and estimated states (K) that takes chance agreement into account and ranges from 0 (no better than chance) to 1 (perfect agreement not attributable to chance). Multiple-imputation pooled estimates are based on $n = 400$ realizations of the position process.

(ESM of Appendix B), and the mean errors in these estimated locations were 81.1, 142.3, and 280.1 m, respectively.

For the four-state simulations ($T = 6519$), each individual HMM typically required about 5 min to converge. All $n = 400$ imputed data analyses required about 12 hr (range = 2–16 h) when run in parallel on a 3.7 GHz processor with eight cores. Both the single- and multiple-imputation approaches generally performed well in all scenarios with respect to state and μ_t estimation, including the ice-associated non-diving states (Fig. 3). The estimated probability distributions for the dive time, dry time, and number of dive data streams were generally unbiased, but as with all other simulations examined here, the estimated step length and turning angle distributions became more biased as measurement error increased and sampling rate decreased (ESM of Appendix B). With high measurement error and low sampling rates, it is clear that the step length and turning angle distributions for the non-diving states ("hauled out on ice" and "resting at sea") would be indistinguishable from the "foraging" state in the absence of additional data streams. While the multiple-imputation approach tended to estimate 95% confidence bands that covered μ_t, mean coverage of μ_t when $M = 50$ m was as low as 84% for $\lambda = 2$ (ESM of Appendix B) with a mean error of 173.7m.

I was somewhat surprised that measurement error and sampling rate did not have a larger impact on identification of the two non-diving states ("hauled out on ice" and "resting at sea") in the four-state simulations. I therefore performed an identical analysis excluding the "dry time" data stream from the fitted HMMs and thus leaving sea ice concentration as the sole data stream distinguishing these two non-diving states. However, I still found that measurement error and sampling rate did not have a deleterious effect on state or location estimation (ESM of Appendix C), and it appears the beta distribution approximation for sea ice used by McClintock et al. (2017) can be useful for identifying different types of resting behavior in ice-associated seals.

For all three sets of simulations, pooled estimates based on all $n = 400$ and a subset of $n = 30$ randomly selected imputations were virtually indistinguishable. However, while pooled estimates based on $n = 5$ imputations generally performed better than the single-imputation approach, they did not perform as well as the pooled estimates from $n = 400$ or $n = 30$ imputations, particularly in the estimation of the true underlying position process (ESM of Appendix B).

4. BEARDED SEAL EXAMPLE

4.1. EXAMPLE METHODS

The bearded seal is a documented benthic forager whose life history is intricately linked with sea ice, but the nature of that relationship and other ecological connections with habitat is poorly understood. McClintock et al. (2017; hereafter MLCB) recently conducted a single-stage analysis of bearded seal biotelemetry data that incorporated six behavior states ("hauled out on ice," "resting at sea," "hauled out on land," "mid-water foraging," "benthic foraging," and "transit"), seven data streams (step length, bearing, proportion dive time, proportion dry time, number of benthic dives, proportion sea ice cover, and proportion
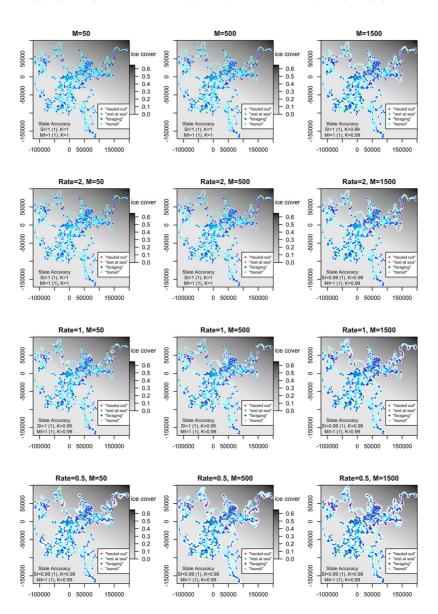
Figure 3. Selected simulation results under the four-state ("hauled out on ice" = *red*, "resting at sea" = *green*, "foraging" = *blue*, and "transit" = *light blue*) scenarios. *Each panel* presents the estimated tracks for $N = 7$ individuals relative to a spatially correlated sea ice concentration grid ("ice cover") under varying levels of location measurement error ($M \in \{50, 500, 1500\}$) and temporally irregular sampling (Rate $\in 2, 1, 0.5$). The first row pertains to scenarios with temporally regular sampling that exactly matches the time steps of the simulated tracks. *Solid lines* indicate the tracks and state sequences using the single-imputation (SI) approach, while white lines indicate $n = 400$ realizations of the position process using the multiple-imputation (MI) approach. Under the data-generating model, both the "hauled out on ice" and "resting at sea" states were associated with short step lengths and high directional persistence, but individuals could only switch to the "hauled out on ice" state if the grid cell of its current position contained >0.05 sea ice concentration. State classification accuracy for each scenario is based on the proportion of time steps for the estimated most likely state sequence that match the true state, the proportion of estimated state probabilities with at least 0.05 probability assigned to the true state (in parentheses), and a measure of agreement between the true and estimated states (K) that takes chance agreement into account and ranges from 0 (no better than chance) to 1 (perfect agreement not attributable to chance). Multiple-imputation pooled estimates are based on $n = 400$ realizations of the position process (Color figure online).

land cover), location measurement error and temporal irregularity, and missing auxiliary biotelemetry data. To formally account for these sources of uncertainty, MLCB formulated their six-state hierarchical HMM as a Bayesian model using a complete data likelihood that conditions on the unobserved states. Their analysis required substantial model development, a custom MCMC model-fitting algorithm written in the C programming language, and several weeks of run time to achieve their target convergence diagnostics. While here I compare the two-stage approach to that MLCB, it deserves note that alternative single-stage model specifications, model-fitting algorithms, or software could lead to decreases (or increases) in computation times relative to MLCB (e.g., Turek et al. 2016).

The bearded seal data consist of $N = 7$ individuals deployed with Argos (Service Argos 2013) satellite-linked tags near Kotzebue, Alaska, USA, between 2009 and 2012. In addition to location data, the tags were equipped with sensors for recording dive and wet/dry data from which the dive time, dry time, and the number of benthic dive data streams were calculated at 6-hr time steps over the duration of each deployment. Given the nature of the Argos platform, bearded seal diving behavior leading to limited or irregular exposure to satellites, and limited bandwidth for transferring data, the location data were subject to measurement error and temporal irregularity, while the auxiliary biotelemetry data were subject to missing or incomplete records. The Argos error ellipses were overwhelmingly oriented toward the x-axis, with mean semi-major axis $M = 11252$ m (median $= 4174$ m, SD $= 36190$), semi-minor axis $m = 493$ m (median $= 239$ m, SD $= 4894$), and orientation $c = 90°$ (median $= 89°$, SD $= 20$). There was an average of 7.2 (SD $= 5.5$) location observations per 6-h time step, with 18% of time steps having no observed locations. Full details of the data can be found in London (2016).

I approximated the model implemented by MLCB using the same single- and multiple-imputation approaches described in *Simulation methods*. The six-state HMM assumes step length $s_t \mid z_t = i \sim \text{Gamma}(a_i, b_i)$, turning angle $\phi_t \mid z_t = i \sim \text{wCauchy}(0, \rho_i)$, and the number of benthic dives $d_t \mid z_t = i \sim \text{Poisson}(r_i)$ for $z \in \{I, S, L, M, B, T\}$, where I denotes "hauled out on ice," S denotes "resting at sea," L denotes "hauled out on land," M denotes "mid-water foraging," B denotes "benthic foraging," T denotes "transit," and wCauchy$(0, \rho_z)$ denotes a wrapped Cauchy distribution with mean zero and concentration parameter $\rho_z \in (-1, 1)$. For the data streams corresponding to proportions ($w_t = \text{dive time}$, $v_t = \text{dry time}$, $c_t = \text{sea ice cover}$, $l_t = \text{land cover}$), the HMM assumes $f_t \mid z_t = i \sim \text{Beta}\left(\upsilon_i^f, \delta_i^f\right)$ for $f \in \{w, v, c, l\}$. To facilitate comparisons and avoid label switching among the n HMM fits, I used the same constraints on the state-dependent probability distribution parameters as MLCB.

Because these tags produced both GPS and Argos locations, the error model for the first-stage "crawl" model fits depended on the location type. For GPS locations, I used a bivariate normal model $y_k \sim N(\mu_k, \Sigma_k)$ assuming $M = 50$. For the Argos locations, I used a bivariate normal model based on the Argos error ellipse (McClintock et al. 2015). For each realization of the position process, the step length, turning angle, sea ice cover, and land cover data streams were calculated at temporally regular time steps matching the 6-hr resolution of the auxiliary biotelemetry data. Unlike the other biotelemetry data streams, the number of benthic dives (d_t), defined as the number of (presumably foraging) dives to the

sea floor, was not directly observable because the exact locations (and thus sea floor depth) during each 6-hr time step were unknown. I therefore calculated the number of benthic dives based on the sea floor depths at the start and end locations for each time step (see MLCB for further details on how benthic dives were calculated). All analyses were performed in R using the "momentuHMM" package.

There are seven key differences between the "full" treatment of MLCB and the two-stage approaches. These differences include: (1) MLCB used MCMC to fit a Bayesian model based on a complete data likelihood that conditions on the unobserved states, while the two-stage approaches maximize the HMM likelihood (Eq. 1) directly; (2) the fitted HMMs using the two-stage approaches are not hierarchical and thus do not contain individual-level random effects on the state-dependent probability distribution parameters; (3) the two-stage approach does not incorporate environmental data (e.g., sea floor depth, sea ice, land) or a maximum bearded seal travelling speed of 2 m/s into the position process; (4) the two-stage approaches used here do not impute any missing auxiliary biotelemetry data (dive time, dry time, and number of benthic dives); (5) the two-stage approaches use a bivariate normal Argos error ellipse model instead of a bivariate t-distributed model because the latter is not implemented in "crawl"; (6) the two-stage approaches utilize a single-state continuous-time correlated random walk model for the position process instead of a multistate discrete-time correlated random walk model (see McClintock et al. 2014); and (7) the two-stage approaches assume the initial distribution (δ) is the stationary distribution. These differences clearly suggest that the results from the analyses will not be identical, but my objective was to assess how well the two-stage approaches approximate the approach of MLCB.

4.2. EXAMPLE RESULTS

With 6 states and $T = 7414$ time steps, each individual HMM required 20–140 mins to converge. Using the single-imputation parameter estimates as initial values for the optimization, all $n = 400$ imputed data analyses required about 70 h when run in parallel on seven cores of a 3.7 GHz processor. Plots of the estimated true locations (μ_t) and states for the single- and multiple-imputation analyses (Fig. 4) both appear qualitatively similar to analogous plots in McClintock et al. (2017). A comparison of the estimated activity budgets based on the Viterbi algorithm (for the two-stage approaches) and posterior state summaries of MLCB indicate these were very similar for the three non-diving states and the "mid-water foraging" states, but some differences were found for the "benthic foraging" and "transit" states (Fig. 5). However, activity budget estimates based on the Viterbi algorithm do not account for uncertainties reflected in the state probabilities ($q_{z,t}$) estimated using the forward–backward algorithm. A closer examination of $q_{z,t}$ for the single- and multiple-imputation approaches indicated that 7% and 3.5% of all time steps, respectively, failed to include $\geq 5\%$ probability for the most likely state identified by MLCB. When they did significantly differ, the state sequences of the two-stage approaches tended to assign the "transit" and "mid-water foraging" states to time steps that were assigned to "benthic foraging" and "transit" by MLCB, respectively.

While most of the estimated probability distributions for each data stream were very similar across all three analyses, the estimated step length, turning angle, and number of benthic
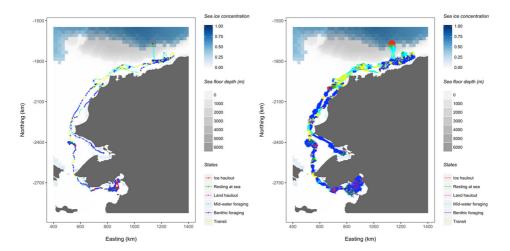
Figure 4. Predicted locations and states for an adult male bearded seal tag deployment from June 25, 2009, to March 9, 2010, near Alaska, USA, using the single-imputation (*left panel*) and multiple-imputation (*right*) approaches. The sea ice concentration in 25 × 25 km grid cells on July 29, 2009, indicates both approaches identified a mid-water foraging trip to the deeper Canada Basin waters off the Beaufort Shelf during which the animal hauled out on the northwardly receding sea ice edge. For the multiple-imputation approach, the results from each of $n = 400$ realizations of the position process are plotted.

dive distributions somewhat differed for the three diving states (Fig. 6, ESM of Appendix D). As the underlying HMMs are very similar, these deviations are likely attributable to the inherent differences between the two-stage approaches and MLCB. Notably, these differences seem most likely due to the smoothness of the tracks generated by the single-state continuous-time correlated random walk model, and the fact that the environmental data (sea floor depth, sea ice, land) are not used to inform the position process in the two-stage approaches. For example, unlike MLCB, the two-stage approaches do not prohibit the position process from moving inland or through waters shallower than the dive depths observed for each time step, thereby allowing for movements close to land where shallower (likely non-foraging) dives would tend to be included in the number of benthic dives for each time step.

5. DISCUSSION

Motivated by my experiences with complex movement HMMs that incorporate measurement error and temporally irregular or missing data, I have investigated the utility of a two-stage approach based on multiple imputation as a more practical alternative to computationally demanding and often time-consuming model-fitting techniques such as MCMC. Based on limited simulations and the bearded seal example, I found this approach to be promising in terms of its ease of implementation, computation time, and performance. The two-stage approach can be performed with existing software using maximum likelihood methods and thus alleviates the need for custom model-fitting algorithms and computer code that is typically required for analogous single-stage analyses. Because maximum likelihood methods are used, likelihood-based model selection criteria (e.g., AIC, BIC) could
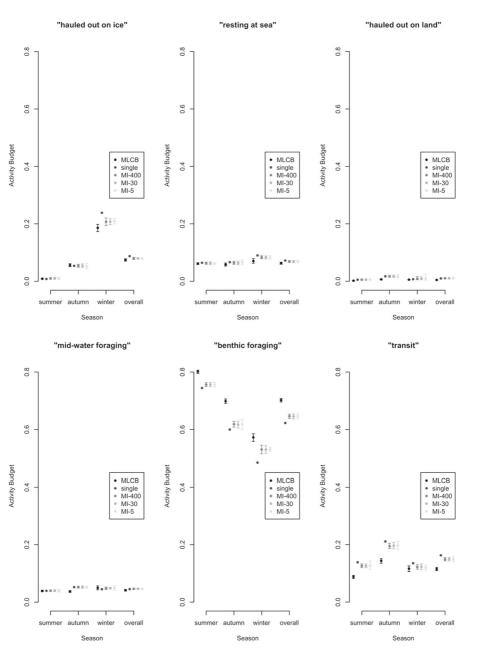
Figure 5. Estimated activity budgets among six behavioral states ("hauled out on ice," "resting at sea," "hauled out on land," "mid-water foraging," "benthic foraging," and "transit") from $N = 7$ bearded seal tag deployments between 2009 and 2012 near Alaska, USA. Results from the single- and multiple-imputation analyses are presented alongside those reported by McClintock et al. (2017; MLCB). There were seasonal differences in activity budgets between "summer" (from tagging in late June and early July to 30 September), "autumn" (1 October to 31 December), and "winter" (1 January until tag loss between February and April) that coincided with the southern advance of winter sea ice in the Arctic. *Error bars* representing 95% confidence and highest posterior density intervals are included for the multiple-imputation and MLCB estimates, respectively. For the multiple-imputation analyses, pooled estimates are based on $n = 400$ ("MI-400"), $n = 30$ ("MI-30"), and $n = 5$ ("MI-5") realizations of the position process.
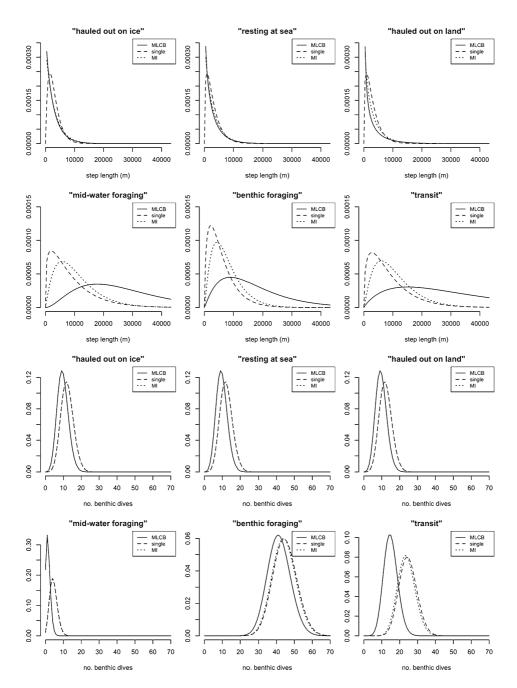
Figure 6. Estimated state-dependent probability distributions for step length (top two rows) and the number of benthic dives (bottom two rows) for $N = 7$ bearded seals using the single-imputation, multiple-imputation (MI), and McClintock et al. (2017; MLCB) approaches. The six behavior states include "hauled out on ice," "resting at sea," "hauled out on land," "mid-water foraging," "benthic foraging," and "transit." For plots in the third row, the single- and multiple-imputation probability densities for the number of benthic dives are nearly identical. Multiple-imputation results are based on $n = 400$ realizations of the position process.

potentially be used to select among competing models at either stage in the analysis. Further, unlike MCMC and similar techniques, multiple imputation is completely parallelizable; with sufficient processing power computation times need not be longer than that required to fit a single HMM. While $n = 400$ imputations were used here, I found that far fewer imputations may actually be necessary in practice. However, the appropriate number of imputations is likely to be case dependent.

In terms of state identification and path reconstruction, I generally found multiple imputation of the position process based on the continuous-time correlated random walk model of Johnson et al. (2008) to be robust to sampling rate and location measurement error. The two-stage approach thus appears to be a reliable method for inferring both when and where individuals exhibit particular behaviors and could therefore be used to investigate hypotheses about activity budgets, space use, resource selection, and many other areas of movement and behavioral ecology. Somewhat to my surprise, low sampling rates and high measurement error did not adversely affect identification of very similar states solely distinguishable by a spatially correlated covariate (i.e., "hauled out on ice" and "resting at sea" in the four-state simulations; see ESM of Appendix C), suggesting the beta approximation for sea ice concentration used by McClintock et al. (2017) can be useful for this purpose. Although not investigated here, it is important to note that difficulties arising from measurement error and temporally irregular or missing data will almost certainly be amplified when state-dependent probability distributions are less distinct than they were in my simulations (Beyer et al. 2013). However, as demonstrated in the four-state simulations, additional data streams can help facilitate accurate state estimation when step length and turning angle distributions are similar among states.

In terms of state transition probabilities and probability distribution estimation for data streams that depend on the position process (e.g., step length, turning angle), multiple imputation based on the model of Johnson et al. (2008) was less robust to sampling rate and, in particular, measurement error. While the estimated effects of the spatial covariate on transition probabilities were all significant except in the most extreme cases, both the single- and multiple-imputation approaches increasingly underestimated the true transition probability coefficients as measurement error increased and sampling rate decreased (see ESM of Appendix B). Clearly, finer-scale spatial relationships can become masked as uncertainty about the position process increases. Based on the simulations examined here, low sampling rates and high measurement error can result in the single-state model of Johnson et al. (2008) smoothing the track toward the more dominant state ("transit" in the two-state simulations without a spatial covariate, "foraging" in the two-state simulations with a spatial covariate, and "foraging" in the four-state simulations; see Fig. 1, ESM of Appendix B). However, the smoothed distributions typically remained distinct enough for the states and spatial covariate effects to be reasonably well inferred in the second stage. Thus, while generally reliable for state and location estimation, if primary interest is in unbiased estimation of distributions for data streams or covariate effects that depend on the position process, I would not recommend using this two-stage approach based on the model of Johnson et al. (2008) unless: (1) measurement errors are small relative to the (state-dependent) scales of movement; and (2) sampling rates are high relative to the time step of interest. Although not explored here, alternative movement models or the inclusion of movement covariates (if available) in the

first stage could help mitigate this smoothing. For example, auxiliary biotelemetry data can be used as covariates in the model of Johnson et al. (2008), but these were not used in the first stage of the bearded seal analysis because missing covariate data are not currently permitted in "crawl."

The single-imputation approach generally performed similarly to multiple imputation when sampling rates were high and measurement error was low. It could therefore potentially serve as a fast and reliable alternative to simple linear imputation or more computationally intensive methods for handling temporally irregular or missing location data under these circumstances. However, as expected, the performance of the single-imputation approach declined with lower sampling rates and higher measurement error.

Although I have found imputation to have several potentially desirable qualities with respect to ease of implementation and computation time, this two-stage approach to accounting for measurement error and missing data in movement HMMs has some important limitations relative to more complicated single-stage alternatives. As highlighted in the bearded seal example, perhaps the most important limitation is the separation of the position process (i.e., the movement model) from the latent behavior states and the environment (e.g., bathymetry, land, sea ice). The latter is perhaps best demonstrated in Fig. 4, where the two-stage approach clearly imputed some inland positions that are rather dubious for an ice-associated seal. The two-stage approach used here also does not include individual-level random effects in the HMMs (Altman 2007; DeRuiter et al. 2016). Extending the two-stage approach to incorporate other types of missing data, random effects, and environmental constraints to movement (e.g., land) is the focus of ongoing research. As with any maximum likelihood analysis, starting values for the optimization are also an important consideration when using the two-stage approach.

Here I have focused on imputation of the position process when location data are subject to measurement error and temporal irregularity, but multiple imputation need not be limited to these scenarios. For example, missing auxiliary biotelemetry data (e.g., dive time, dry time, number of dives) in the bearded seal example could be imputed using standard missing data techniques (Rubin 1987). This would also allow for the investigation of different mechanisms for missingness that can be problematic if not accounted for (Nakagawa and Freckleton 2008).

Despite the subtle but important differences between two-stage approaches and the single-stage treatment, the overall inferences from the bearded seal example are quite similar in terms of the primary objectives of McClintock et al. (2017), including the quantification of seasonal activity budgets, the identification of bearded seal foraging habitat, and the characterization of different movement behaviors in relation to seasonal sea ice. However, if one prefers single-stage approaches, the two-stage approach could still prove useful for exploring alternative HMMs from which to choose a potential candidate for the single-stage treatment. Given its ease of implementation and relatively fast computation, the two-stage approach could also be used to generate initial values or help diagnose convergence in single-stage analyses. Whether used for primary analysis or as an exploratory aid, I found multiple imputation to be a promising addition to the ecologist's toolbox for inference when HMM data streams are subject to observation error.

ACKNOWLEDGEMENTS

REFERENCES

Altman, R. M. (2007), "Mixed hidden Markov models", *Journal of the American Statistical Association*, 102, 201–210

Bagniewska, J. M., Hart, T., Harrington, L. A., and Macdonald, D. W. (2013), "Hidden Markov analysis describes dive patterns in semiaquatic animals", *Behavioral Ecology*, 24, 659–667.

Beyer, H. L., Morales, J. M., Murray, D., and Fortin, M. J. (2013), "The effectiveness of Bayesian state-space models for estimating behavioural states from movement paths", *Methods in Ecology and Evolution*, 4, 433–441.

Congalton, R.G. (1991), "A review of assessing the accuracy of classifications of remotely sensed data", *Remote Sensing of Environment*, 37, 35–46.

Costa, D. P., Robinson, P. Arnould, J., Harrison, A-L., Simmons, S., et al. (2010), "Accuracy of ARGOS locations of pinnipeds at-sea estimated using Fastloc GPS", *PLoS ONE*, 5, e8677.

DeRuiter, S. L., Langrock, R., Skirbutas, T., Goldbogen, J. A., Chalambokidis, J., Friedlaender, A. S., and Southall, B. L. (2016), "A multivariate mixed hidden Markov model to analyze blue whale diving behaviour during controlled sound exposures", arXiv:1602.06570.

Hooten, M. B., Johnson, D. S., McClintock, B. T., and Morales, J. M. (2017), *Animal Movement: Statistical Models for Telemetry Data*. CRC Press. Boca Raton, Florida, USA.

Johnson, D. S. (2016), "crawl: Fit Continuous-Time Correlated Random Walk Models to Animal Movement Data", *R package version 2.0.4.* https://cran.r-project.org/web/packages/crawl/

Johnson, D. S., London, J. M., Lea, M. A., and Durban, J. W. (2008), "Continuous-time correlated random walk model for animal telemetry data", *Ecology*, 89, 1208–1215.

Jonsen, I. D., Flemming, J. M. and Myers, R. A. (2005), "Robust state-space modeling of animal movement data", *Ecology*, 86, 2874–2880.

Langrock, R., King, R., Matthiopoulos, J., Thomas, L., Fortin, D., and Morales, J. M. (2012), "Flexible and practical modeling of animal telemetry data: hidden Markov models and extensions", *Ecology*, 93, 2336–2342.

London, J. M. (2016), kotzeb0912: v1.0 [Data set]. *Zenodo*. http://doi.org/10.5281/zenodo.57290

McClintock, B. T., King, R., Thomas, L., Matthiopoulos, J., McConnell, B. J., and Morales, J. M. (2012), "A general discrete-time modeling framework for animal movement using multistate random walks", *Ecological Monographs*, 82, 335–349.

McClintock, B. T., Russell, D. J., Matthiopoulos, J., and King, R. (2013), "Combining individual animal movement and ancillary biotelemetry data to investigate population-level activity budgets", *Ecology*, 94, 838–849.

McClintock, B. T., Johnson, D. S., Hooten, M. B., Ver Hoef, J. M., and Morales, J. M. (2014), "When to be discrete: the importance of time formulation in understanding animal movement", *Movement Ecology*, 2, 21.

McClintock, B. T., London, J. M., Cameron, M. F., and Boveng, P. L. (2015), "Modelling animal movement using the Argos satellite telemetry location error ellipse", *Methods in Ecology and Evolution*, 6, 266–277.

McClintock, B. T., London, J. M., Cameron, M. F., and Boveng, P. L. (2017), "Bridging the gaps in animal movement: hidden behaviors and ecological relationships revealed by integrated data stream", *Ecosphere*, 8, e01751.

Michelot, T., Langrock, R., and Patterson, T. A. (2016a), "moveHMM: an R package for the statistical modelling of animal movement data using hidden Markov models", *Methods in Ecology and Evolution*, 7, 1308–1315.

Michelot, T., Langrock, R., Bestley, S., Jonsen, I. D., Photopoulou, T., and Patterson, T. A. (2016b), "Estimation and simulation of foraging trips in land-based marine predators", arXiv:1610.06953.

Morales, J. M., Haydon, D. T., Frair, J., Holsinger, K. E., and Fryxell, J. M. (2004), "Extracting more out of relocation data: building movement models as mixtures of random walks", *Ecology*, 85, 2436–2445.

Nakagawa, S., and Freckleton, R. (2008), "Missing inaction: the dangers of ignoring missing data", *Trends in Ecology and Evolution*, 23, 592–596.

Pebesma, E. J. (2004), "Multivariable geostatistics in S: the gstat package", *Computers and Geosciences*, 30, 683–691.

R Core Team (2016), *R: a Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. Vienna, Austria. http://www.R-project.org

Rubin, D. (1987), *Multiple Imputation for Nonresponse in Surveys*. Wiley, New York, USA.

Rubin, D. B. and Schenker, N. (1986), "Multiple imputation for interval estimation from simple random samples with ignorable nonresponse", *Journal of the American Statistical Association*, 81, 366–374.

Russell, D. J., McClintock, B., Matthiopoulos, J., Thompson, P., Thompson, D., Hammond, P. Jones, E., MacKenzie, M., Moss, S., and McConnell, B. (2015), "Intrinsic and extrinsic drivers of activity budgets in sympatric grey and harbour seals", *Oikos*, 124, 1462–1472.

Service Argos (2013), *Argos User's Manual*. CLS. http://www.argos-system.org/manual.

Silva, M. A., Jonsen, I., Russell, D. J. F., Prieto, R., Thompson, D., and Baumgartner, M. F. (2014), "Assessing performance of Bayesian state-space models fit to Argos satellite telemetry locations processed with Kalman filtering", *PLoS ONE*, 9, e92277.

Turek, D., de Valpine, P., and Paciorek, C. J. (2016), "Efficient Markov chain Monte Carlo sampling for hierarchical hidden Markov models", *Environmental and Ecological Statistics* 23, 549–564.

Zucchini, W., MacDonald, I. L., and Langrock, R. (2016), *Hidden Markov Models for Time Series: An Introduction Using R*, Second Edition. CRC Press. Boca Raton, Florida, USA.