This is a repository copy of *momentuHMM: R package for generalized hidden Markov models of animal movement*.

momentuHMM: R package for generalized hidden Markov models of animal movement

Brett T. McClintock[1] and Théo Michelot[2]

[1]Marine Mammal Laboratory
Alaska Fisheries Science Center
NOAA National Marine Fisheries Service
Seattle, U.S.A.
Email: brett.mcclintock@noaa.gov

[2]School of Mathematics and Statistics
University of Sheffield
Sheffield, U.K.

RUNNING HEAD: R package momentuHMM

March 13, 2018

Summary

1. Discrete-time hidden Markov models (HMMs) have become an immensely popular tool for inferring latent animal behaviours from telemetry data. While movement HMMs typically rely solely on location data (e.g. step length and turning angle), auxiliary biotelemetry and environmental data are powerful and readily-available resources for incorporating much more ecological and behavioural realism. However, complex movement or observation process models often necessitate custom and computationally-demanding HMM model-fitting techniques that are impractical for most practitioners, and there is a paucity of generalized user-friendly software available for implementing multivariate HMMs of animal movement.

2. Here we introduce an open-source R package, `momentuHMM`, that addresses many of the deficiencies in existing HMM software. Features include: 1) data pre-processing and visualization; 2) user-specified probability distributions for an unlimited number of data streams and latent behaviour states; 3) biased and correlated random walk movement models, including dynamic "activity centres" associated with attractive or repulsive forces; 4) user-specified design matrices and constraints for covariate modelling of parameters using formulas familiar to most R users; 5) multiple imputation methods that account for measurement error and temporally-irregular or missing data; 6) seamless integration of spatio-temporal covariate raster data; 7) cosinor and spline models for cyclical and other complicated patterns; 8) model checking and selection; and 9) simulation.

3. After providing an overview of the main features of the package, we demonstrate some of the capabilities of `momentuHMM` using real-world examples. These include models for cyclical movement patterns of African elephants, foraging trips of northern fur seals, loggerhead turtle movements relative to ocean surface currents, and grey seal movements among three activity centres.

4. `momentuHMM` considerably extends the capabilities of existing HMM software while accounting for common challenges associated with telemetry data. It therefore facilitates more realistic hypothesis-driven animal movement analyses that have hitherto been largely inaccessible to non-statisticians. While motivated by telemetry data, the package can be used for analyzing any type of data that is amenable to HMMs. Practitioners interested in additional features are encouraged to contact the authors.

Key-words biologging, biotelemetry, `crawl`, `moveHMM`, state-space model, state-switching

1 Introduction

Animal movement is central to many ecological processes and considered essential to our understanding of animal behaviour, population dynamics, and the impacts of global change. Coupled with advances in animal-borne biologging technology, there has been a recent explosion in analytical methods for inferring animal movement, behaviour, space use, and resource selection from telemetry data (Hooten *et al.* 2017). Largely attributable to their ease of implementation and interpretation, discrete-time hidden Markov models (HMMs) have become immensely popular for characterizing animal movement behaviour (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 a (possibly multivariate) observation process $(\mathbf{Z}_1, \ldots, \mathbf{Z}_T)$, in which each data stream is generated by N state-dependent probability distributions, and where the unobservable (hidden) state sequence $(S_t \in \{1, \ldots, N\}, t = 1, \ldots, T)$ is assumed to be a Markov chain. The state sequence of the Markov chain is governed by (typically first-order) state transition probabilities, $\gamma_{ij}^{(t)} = \Pr(S_{t+1} = j \mid S_t = i)$ for $i, j = 1, \ldots, N$, and an initial distribution $\boldsymbol{\delta}^{(0)}$. The likelihood of an HMM can be succinctly expressed using the forward algorithm:

$$\mathcal{L} = \boldsymbol{\delta}^{(0)} \boldsymbol{\Gamma}^{(1)} \mathbf{P}(\mathbf{z}_1) \boldsymbol{\Gamma}^{(2)} \mathbf{P}(\mathbf{z}_2) \boldsymbol{\Gamma}^{(3)} \cdots \boldsymbol{\Gamma}^{(T-1)} \mathbf{P}(\mathbf{z}_{T-1}) \boldsymbol{\Gamma}^{(T)} \mathbf{P}(\mathbf{z}_T) \mathbf{1}^N, \qquad (1)$$

where $\boldsymbol{\Gamma}^{(t)} = \left(\gamma_{ij}^{(t)} \right)$ is the $N \times N$ transition probability matrix, $\mathbf{P}(\mathbf{z}_t) = \mathrm{diag}(p_1(\mathbf{z}_t), \ldots, p_N(\mathbf{z}_t))$, $p_s(\mathbf{z}_t)$ is the conditional probability density of $\mathbf{Z}_t$ given $S_t = s$, and $\mathbf{1}^N$ is a N-vector of ones. For a thorough introduction to HMMs, see Zucchini *et al.* (2016).

The most common discrete-time animal movement HMMs for telemetry location data are composed of two data streams (e.g. step length and turning angle) and $N = 2$ hidden states (Morales *et al.* 2004; Jonsen *et al.* 2005). These states are often considered as proxies for animal behaviour and characterized by area-restricted-search-type movements ("encamped" or "foraging") and migratory-type movements ("exploratory" or "transit"). Some common probability distributions for the step length data stream are the gamma or Weibull distributions, while the wrapped Cauchy or von Mises distributions are often employed for the turning angle (e.g. Langrock *et al.* 2012). While HMMs based solely on location data are somewhat limited in the number and type of

biologically-meaningful movement behaviour states they are able to accurately identify, auxiliary biotelemetry and environmental data allow for multivariate HMMs that can incorporate much more behavioural realism and facilitate inferences about complex ecological relationships that would otherwise be difficult or impossible to infer from location data alone (e.g. McClintock *et al.* 2013, 2017; DeRuiter *et al.* 2017).

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 (Eq. 1) and computation of the most likely sequence of hidden states using the Viterbi algorithm (Zucchini *et al.* 2016). In movement HMMs, the decoded state sequence can be useful for determining when and where changes in behaviour occur, identifying habitats associated with particular behaviours (e.g. foraging "hotspots"), and calculating the proportion of time steps allocated to each state (i.e. "activity budgets"). However, location measurement error is rarely non-existent in animal-borne telemetry studies and depends on both the device and the system under study (e.g. Costa *et al.* 2010). Furthermore, telemetry data are often obtained with little or no temporal regularity, as in many marine mammal telemetry studies (e.g. Jonsen *et al.* 2005), such that observations do not align with the regular time steps required by discrete-time HMMs. 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 (e.g. Jonsen *et al.* 2005; McClintock *et al.* 2012). Unfortunately, complex analyses requiring novel statistical methods and custom model-fitting algorithms are not practical for many practitioners.

While statisticians have been applying HMMs to telemetry data for decades, R (R Core Team 2017) packages such as `bsam` (Jonsen *et al.* 2005), `moveHMM` (Michelot *et al.* 2016), and `swim` (Whoriskey *et al.* 2017) have recently helped make these models more accessible to the practitioners that are actually conducting telemetry studies. While these contributions represent important steps forward, the models that can currently be implemented remain limited in many key respects. For example, most HMM software for animal movement is limited to correlated random walks with $N = 2$ states and two data streams (e.g. step length and turning angle). While `moveHMM` does permit $N > 2$, it is typically difficult to identify >2 biologically-meaningful behaviour states from only

two data streams (e.g. Morales *et al.* 2004; Beyer *et al.* 2013; McClintock *et al.* 2014). Both `moveHMM` and `swim` require temporally-regular location data with negligible measurement error. Other notable deficiencies of existing software include limited abilities to incorporate spatio-temporal environmental or individual covariates on parameters, biased movements in response to attractive, repulsive, or environmental forces (e.g. McClintock *et al.* 2012), group dynamic models (Langrock *et al.* 2014), cyclical (e.g. daily, seasonal) and other more complicated behavioural patterns, or constraints on parameters.

To address these deficiencies in existing software, we introduce a new user-friendly R package, `momentuHMM`, intended for practitioners wishing to implement more flexible and realistic HMM analyses of animal movement while accounting for common challenges associated with telemetry data. We provide a brief overview of `momentuHMM` and demonstrate some of its capabilities using real-world examples. Step-by-step tutorials, help files, examples, and code can be found in the package's documentation and vignette. This article describes `momentuHMM` version 1.4.0.

2   momentuHMM overview

The workflow for `momentuHMM` analyses consists of several steps (Fig. 1). These include data preparation, data visualization, model specification and fitting, results visualization, and model checking. The workhorse functions of the package are listed in Table 1. Usage of several of these functions is deliberately similar to equivalent functions in `moveHMM` (Michelot *et al.* 2016), but the `momentuHMM` arguments have been generalized and expanded to accommodate a much more flexible framework for data pre-processing, model specification, parameterization, visualization, and simulation. R users already familiar with `moveHMM` will therefore likely find it easy to immediately begin using `momentuHMM`.

A key feature of `momentuHMM` is the ability to include an unlimited number of HMM data streams (e.g. step length, turning angle, dive activity, heart rate) arising from a broad range of commonly used probability distributions (e.g. beta, gamma, normal, Poisson, von Mises, Weibull). Using model formulas familiar to most R users, any of the data stream probability distribution parameters can be modelled as functions of environmental and individual covariates via link functions (e.g. McCullagh & Nelder 1989,
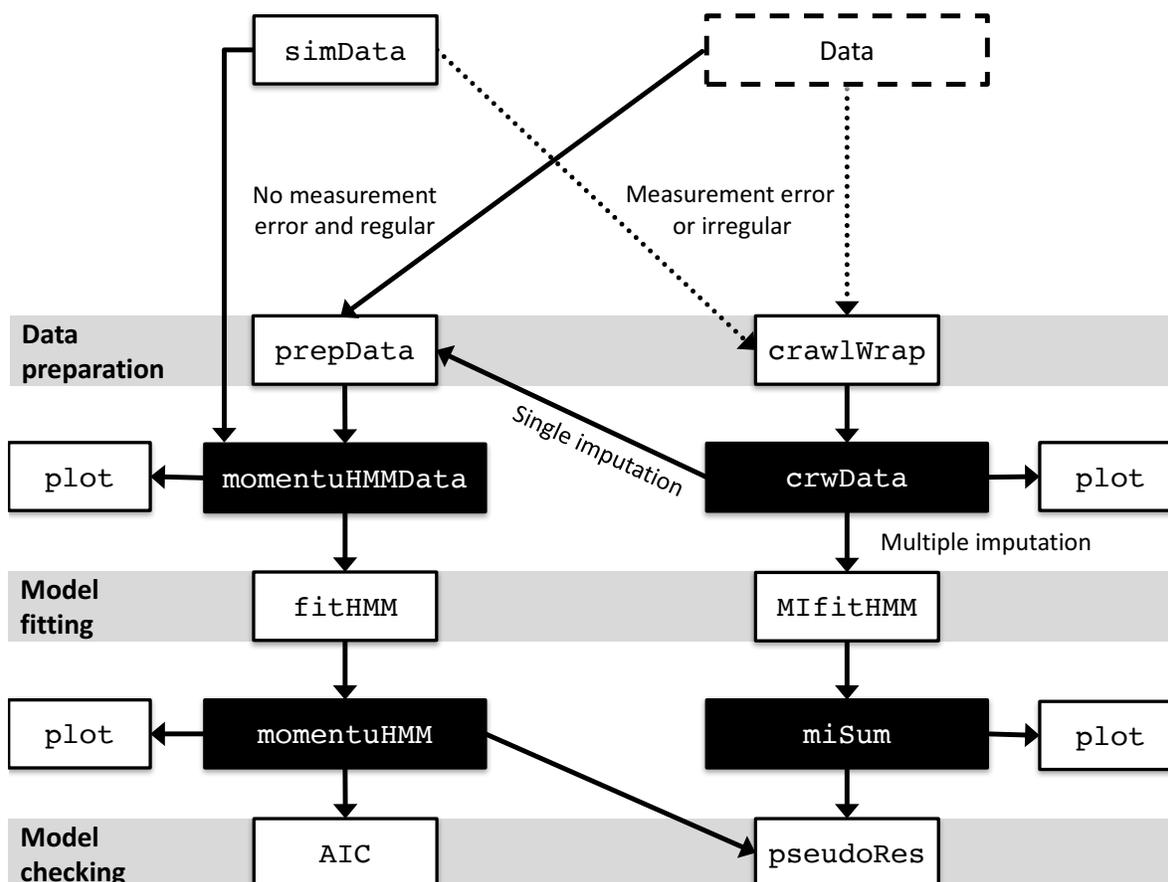
Figure 1. *Schematic representing the typical momentuHMM workflow. White boxes indicate package functions and black boxes indicate object classes returned by functions. The simData function can be used to simulate multivariate HMM data from scratch or from a fitted model. When telemetry data are subject to location measurement error or temporal irregularity (dotted arrows), these sources of uncertainty can be incorporated into HMMs using single or multiple imputation of the position process via the crawlWrap function.*

Table 1. *Workhorse functions for the R package momentuHMM.*

Function	Description
crawlMerge	Merge `crawlWrap` output with additional data streams or covariates
crawlWrap	Fit continuous-time correlated random walk (CTCRW) models and predict temporally-regular locations using the `crawl` package
fitHMM	Fit a (multivariate) HMM to the data
MIfitHMM	Fit (multivariate) HMMs to multiple imputation data
MIpool	Pool `momentuHMM` model results across multiple imputations
plot.crwData	Plot `crawlWrap` output
plot.miSum	Plot summaries of multiple imputation `momentuHMM` models
plot.momentuHMM	Plot summaries of `momentuHMM` models
plot.momentuHMMData	Plot summaries of selected data streams and covariates
plotPR	Plot time series, qq-plots and sample ACFs of pseudo-residuals
plotSat	Plot locations on satellite image
plotSpatialCov	Plot locations on raster image
plotStates	Plot the (Viterbi-)decoded states and state probabilities
prepData	Pre-process data streams and covariates
pseudoRes	Calculate pseudo-residuals for `momentuHMM` models
simData	Simulate data from a (multivariate) HMM
stateProbs	State probabilities for each time step
viterbi	Most likely state sequence (using the Viterbi algorithm)

and see Table 2). The initial distribution ($\boldsymbol{\delta}^{(0)}$) and state transition probabilities ($\gamma_{ij}^{(t)}$) can also be modelled as functions of covariates. Permissable R classes for covariates include `numeric`, `integer`, or `factor`. Spatio-temporal covariates (e.g. wind velocity, forest cover, sea ice concentration) can also be `raster` classes (Hijmans 2016), in which case `momentuHMM` extracts the appropriate values based on the time and location of each observation (see section 3.3).

2.1 Data preparation and visualization

For temporally-regular location data with negligible measurement error, the `prepData` function returns a `momentuHMMData` object (i.e. a data frame of class `momentuHMMData`) that can be used for data visualization and further analysis. Summary plots of `momentuHMMData` objects can be created for any data stream or covariate using the generic `plot` function. If location data are temporally-irregular or subject to measurement error, then a 2-stage multiple imputation approach can be performed (Hooten *et al.* 2017; McClintock 2017). We discuss this pragmatic approach to incorporating location uncertainty into HMMs of animal movement in section 2.3.

2.2 HMM specification and fitting

The function `fitHMM` is used to specify and fit a (multivariate) HMM using maximum likelihood methods. There are many different options for `fitHMM`, so here we will only focus on several of the most important and useful features. The bare essentials of `fitHMM` include the arguments:

- `data` A `momentuHMMData` object

- `nbStates` Number of latent states (N)

- `dist` A list indicating the probability distributions for the data streams

- `estAngleMean` A list indicating whether or not to estimate the mean for angular data streams (e.g. turning angle)

- `formula` Model formula for state transition probabilities ($\gamma_{ij}^{(t)}$)

- `stationary` Logical indicating whether or not the initial distribution ($\boldsymbol{\delta}^{(0)}$) is considered equal to the stationary distribution

Table 2. *Data stream (z) probability distributions, natural parameters, and default link functions for covariate modelling. Probability distributions with positive support can be zero-inflated (with additional zero-mass parameters), while the beta distribution can be zero- and one-inflated (with additional one-mass parameters). If user-specified bounds are provided, then custom link functions are used instead of the defaults (see momentuHMM package documentation for further details). If circular-circular regression is specified for the mean of angular distributions ("vm" and "wrpcauchy"), then a link function based on Rivest et al. (2016) is used. The von Mises consensus distribution ("vmConsensus") is a von Mises circular-circular regression model where the concentration parameter depends on the level of agreement among short-term directional persistence and angular covariates. Users seeking additional data stream probability distributions are encouraged to contact the authors.*

Distribution	Support	Parameters	Link function[1]
Bernoulli ("`bern`")	$z_t \in \{0, 1\}$	`prob` $\in (0, 1)$	logit
Beta ("`beta`")	$z_t \in (0, 1)$	`shape1` > 0	log
		`shape2` > 0	log
		`zero-mass` $\in (0, 1)$	logit
		`one-mass` $\in (0, 1)$	logit
Exponential ("`exp`")	$z_t > 0$	`rate` > 0	log
		`zero-mass` $\in (0, 1)$	logit
Gamma ("`gamma`")	$z_t > 0$	`mean` > 0	log
		`sd` > 0	log
		`zero-mass` $\in (0, 1)$	logit
Log normal ("`lnorm`")	$z_t > 0$	`location` $\in \mathbb{R}$	identity
		`scale` > 0	log
		`zero-mass` $\in (0, 1)$	logit
Normal ("`norm`")	$z_t \in \mathbb{R}$	`mean` $\in \mathbb{R}$	identity
		`sd` > 0	log
Poisson ("`pois`")	$z_t \in \{0, 1, \ldots\}$	`lambda` > 0	log
Von Mises ("`vm`")	$z_t \in (-\pi, \pi]$	`mean` $\in (-\pi, \pi]$	$\tan($`mean`$/2)$
		`concentration` > 0	log
Von Mises ("`vmConsensus`")	$z_t \in (-\pi, \pi]$	`mean` $\in (-\pi, \pi]$	Rivest *et al.* (2016)
		`kappa` > 0	log
Weibull ("`weibull`")	$z_t > 0$	`shape` > 0	log
		`scale` > 0	log
		`zero-mass` $\in (0, 1)$	logit
Wrapped Cauchy ("`wrpcauchy`")	$z_t \in (-\pi, \pi]$	`mean` $\in (-\pi, \pi]$	$\tan($`mean`$/2)$
		`concentration` $\in (0, 1)$	logit

[1]Link functions (g) relate natural scale parameters ($\boldsymbol{\theta}$) to a $T \times k$ design matrix ($\mathbf{X}$) and $k-$vector of working scale parameters ($\boldsymbol{\beta} \in \mathbb{R}^k$) such that $g(\boldsymbol{\theta}) = \mathbf{X}\boldsymbol{\beta}$.

- `Par0` A list of starting values for the probability distribution parameters of each data stream

These seven arguments are all that are needed in order to fit the HMMs currently supported in `moveHMM` (Michelot *et al.* 2016), but the similarities between the packages largely end here. Many of the arguments in `fitHMM` and other `momentuHMM` functions are lists, with each element of the list corresponding to a data stream. Additional data streams can be added to HMMs by simply adding additional elements to these list arguments (see section 3.2).

The `formula` argument can include many of the functions and operators commonly used to construct terms in R model formulas (e.g. `a*b`, `a:b`, `cos(a)`). It can also be used to specify transition probability matrix models that incorporate cyclical patterns using the `cosinor` function (see section 3.1), splines for explaining other more complicated patterns (e.g. `bs` and `ns` functions in the R base package `splines`), and factor variables (e.g. group- or individual-level effects). By default the `formula` argument applies to all entries in the transition probability matrix ($\mathbf{\Gamma}^{(t)}$), but special functions allow for state- and parameter-specific formulas to be specified. Specific state transition probabilities can also be fixed to zero (or any other value), which can be useful for incorporating more behavioural realism by prohibiting or enforcing switching from one particular state to another (possibly as a function of spatio-temporal covariates). Similar arguments are available for initial distribution parameters.

The `DM` argument of `fitHMM` allows covariate models to be specified for the state-dependent probability distribution parameters of each data stream. `DM` formulas are just as flexible as the `formula` argument. In lieu of formulas, `DM` can also be specified as a "pseudo-design" matrix with rows and columns corresponding to the natural and working scale parameters, respectively (Table 2). Pseudo-design matrices allow working parameters to be shared among natural parameters for imposing relational constraints (e.g. $\mu_1 = \mu_2$, $\mu_1 \leq \mu_2$). Probability distribution parameters can also be fixed to user-specified values.

Another noteworthy `fitHMM` argument, `circularAngleMean`, enables circular-circular regression models for the mean direction of angular distributions, such as the wrapped Cauchy and von Mises, instead of circular-linear models based on the tangent link function (Table 2). As a trade-off between biased and correlated movements, circular-circular regression models can weight the attractive (or repulsive) strengths of angular

covariates relative to short-term directional persistence using a special link function based on Rivest *et al.* (2016). Many interesting hypotheses can be addressed using circular-circular regression, including the effects of wind, sea surface currents, and centres of attraction on animal movement direction (see sections 3.3 and 3.4).

2.3 Multiple imputation

When location data are temporally-irregular or subject to measurement error, `momentuHMM` can be used to perform the 2-stage multiple imputation approach of McClintock (2017). There are three primary functions for performing multiple imputation HMM analyses: `crawlWrap`, `MIfitHMM`, and `MIpool`. Based on the R package `crawl` (Johnson 2017), `crawlWrap` is a wrapper function for fitting the continuous-time correlated random walk (CTCRW) model of Johnson *et al.* (2008) and then predicting temporally-regular tracks of the user's choosing (e.g. 15 min, hourly, daily). `MIfitHMM` is essentially a wrapper function for `fitHMM` that repeatedly fits the same user-specified HMM to m imputed data sets using maximum likelihood. Based on the m model fits, the `MIpool` function calculates pooled estimates, standard errors, and confidence intervals using standard multiple imputation formulae. 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 m realizations of the position process, and the pooled inferences across the fitted HMMs therefore reflect location uncertainty.

2.4 Model visualization and diagnostics

The generic `plot` function for `momentuHMM` model objects plots the data stream histograms along with their corresponding estimated probability distributions, the estimated natural parameters and state transition probabilities as a function of any covariates included in the model, and the tracks of all individuals (colour-coded by the most likely state sequence). Confidence intervals for the natural parameters and state transition probabilities can also be plotted. For multiple imputation analyses, estimated 95% location error ellipses can be included in plots of individual tracks. Diagnostic tools include the calculation and plotting of pseudo-residuals (Zucchini *et al.* 2016) using the `pseudoRes` and `plotPR` functions, respectively. Akaike's Information Criterion can be calculated for one or more models using the `AIC.momentuHMM` function.

2.5 Simulation

The function `simData` can be used to simulate multivariate HMMs from scratch or from a fitted model. The arguments are very similar to those used for data preparation and model specification, but they include additonal arguments for simulating location data subject to temporal irregularity and measurement error. Among other features, activity centres and rasters of spatio-temporal covariates can be utilized (see section 3.4). `simData` can therefore simulate more ecologically-realistic tracks (potentially subject to observation error) for study design, power analyses, and assessing model performance or goodness-of-fit (Morales *et al.* 2004).

3 Examples

We now demonstrate some of the capabilities of `momentuHMM` using real telemetry data. Intended for demonstration purposes only, we do not claim these represent improvements relative to previous or alternative analyses for these data sets. While space limits us to only a few key elements here, complete details and code for reproducing all examples described herein are provided in the "vignettes" source directory.

3.1 African elephant

As our first example, we use an African elephant (*Loxodonta africana*) bull track described in Wall *et al.* (2014). In addition to hourly locations, the tag collected external temperature data. Using `fitHMM`, we fitted a 2-state HMM assuming a wrapped Cauchy distribution for turning angle and a gamma distribution for step length. Autocorrelation function estimates suggest there are 24 h cycles in the step length data (Fig. 2), and this presents an opportunity to demonstrate the cosinor model for incorporating cyclical behaviour in model parameters. We therefore specified temperature effects on the turning angle concentration parameters and cycling temperature effects (with a 24 h periodicity) on the step length and state transition probability parameters using the `DM` and `formula` arguments.

The model identifed a state of slow undirected movement ("encamped"), and a state of faster and more directed movement ("exploratory") (Fig. 2). Interestingly, this model suggests step lengths and directional persistence for the "encamped" state decreased as temperature increased, step lengths for both states tended to decrease in
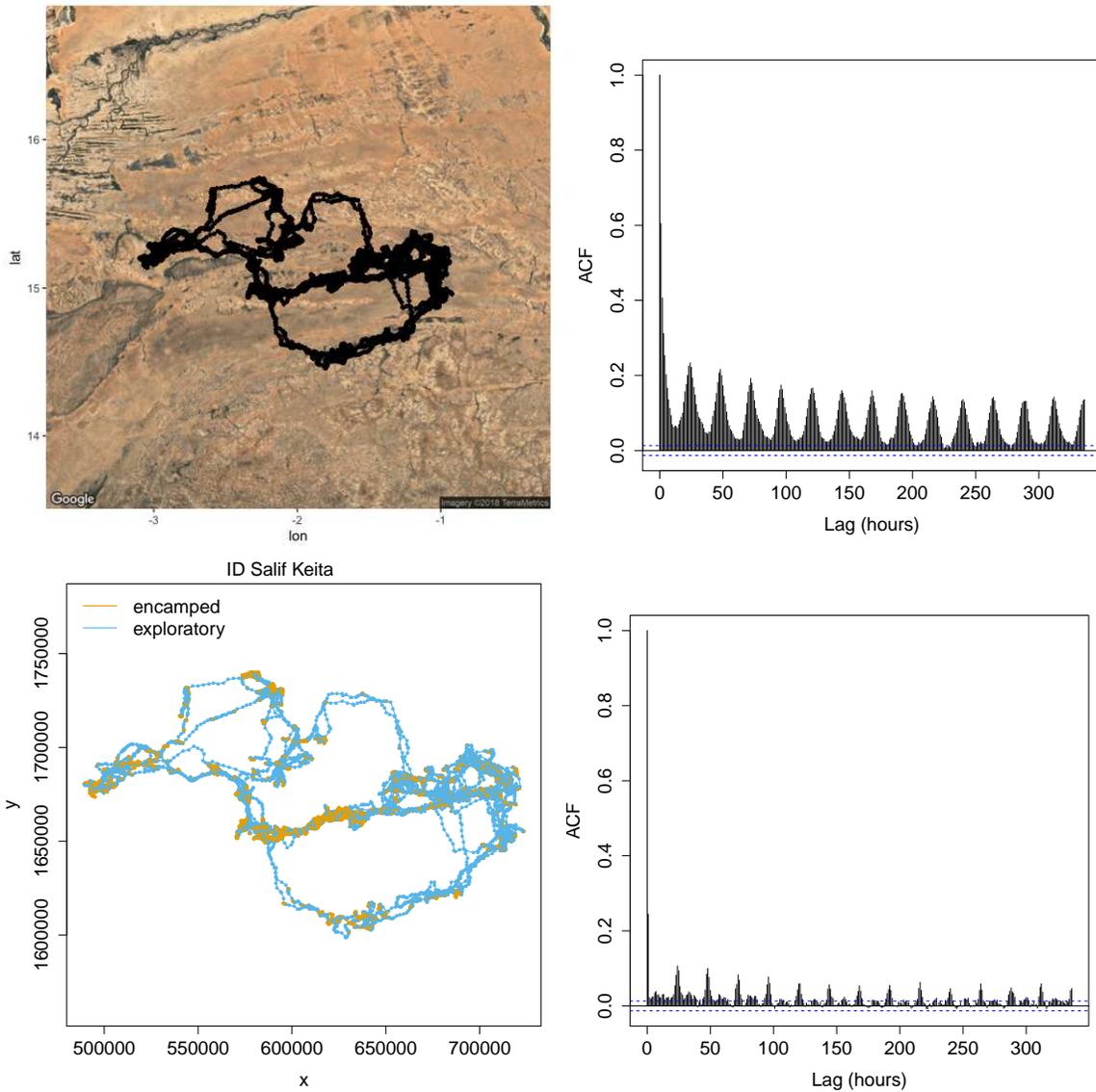
Figure 2. *Selected plots for the 2-state ("encamped" and "exploratory") African elephant example. About 74% and 26% of the 1 h time steps were attributed to the "encamped" and "exploratory" states, respectively. Top-left panel is a plot of the elephant track produced using the 'plotSat' function, top-right panel presents the autocorrelation function (ACF) plot of the step length data, bottom-left panel is a plot of the Viterbi-decoded track (orange = "encamped", blue = "exploratory") generated from the fitted model using the generic 'plot' function, and bottom-right panel presents the step length pseudo-residual ACF plot using the 'plotPR' function.*

the late evening and early morning, and transition probabilities from the "encamped" to "exploratory" state decreased as temperature increased (Fig. 3). This model was overwhelmingly supported by AIC when compared to alternative models with fewer covariates, but model fit can be further assessed using pseudo-residuals. In this case, the pseudo-residuals indicate the model explained much of the periodicity in step length, although there does still appear to be some room for improvement (Fig. 2).

3.2   Northern fur seal

We use the northern fur seal (*Callorhinus ursinus*) example from McClintock *et al.* (2014) to demonstrate the use of additional data streams for distinguishing behaviours with similar horizontal trajectories (e.g. "resting" and "foraging") in a multivariate HMM. The data consist of 241 temporally-irregular Fastloc GPS locations obtained during a foraging trip of a nursing female near the Pribilof Islands of Alaska, USA. The tag included time-depth recording capabilities, and the dive activity data were summarized as the number of foraging dives over $T = 228$ temporally-regular 1 h time steps. To fit the $N = 3$ state (1="resting", 2="foraging", 3="transit") model of McClintock *et al.* (2014), we first used `crawlWrap` to predict temporally-regular locations at 1 h time steps assuming a bivariate normal location measurement error model. We then used multiple imputation to account for location uncertainty by repeatedly fitting the 3-state HMM to $m = 100$ realizations of the position process using `MIfitHMM`. We specified a gamma distribution for step length, wrapped Cauchy distribution for turning angle, and Poisson distribution for the number of foraging dives.

Our results were very similar to those of the discrete-time Bayesian model of McClintock *et al.* (2014), with periods of "foraging" often followed by "resting" at sea (Fig. 4). Activity budgets based on the estimated state sequences for each imputation were 0.31 (95% CI: 0.23−0.4) for "resting", 0.28 (95% CI: 0.23−0.34) for "foraging", and 0.41 (95% CI: 0.31−0.51) for "transit".

3.3   Loggerhead turtle

Using hitherto unpublished loggerhead turtle (*Caretta caretta*) data for a captive-raised juvenile released on the coast of North Carolina, USA, we now demonstrate how movement direction and step length can be easily modelled as a function of angular covariates in `momentuHMM`. The data consist of temporally-irregular Argos locations and rasters of
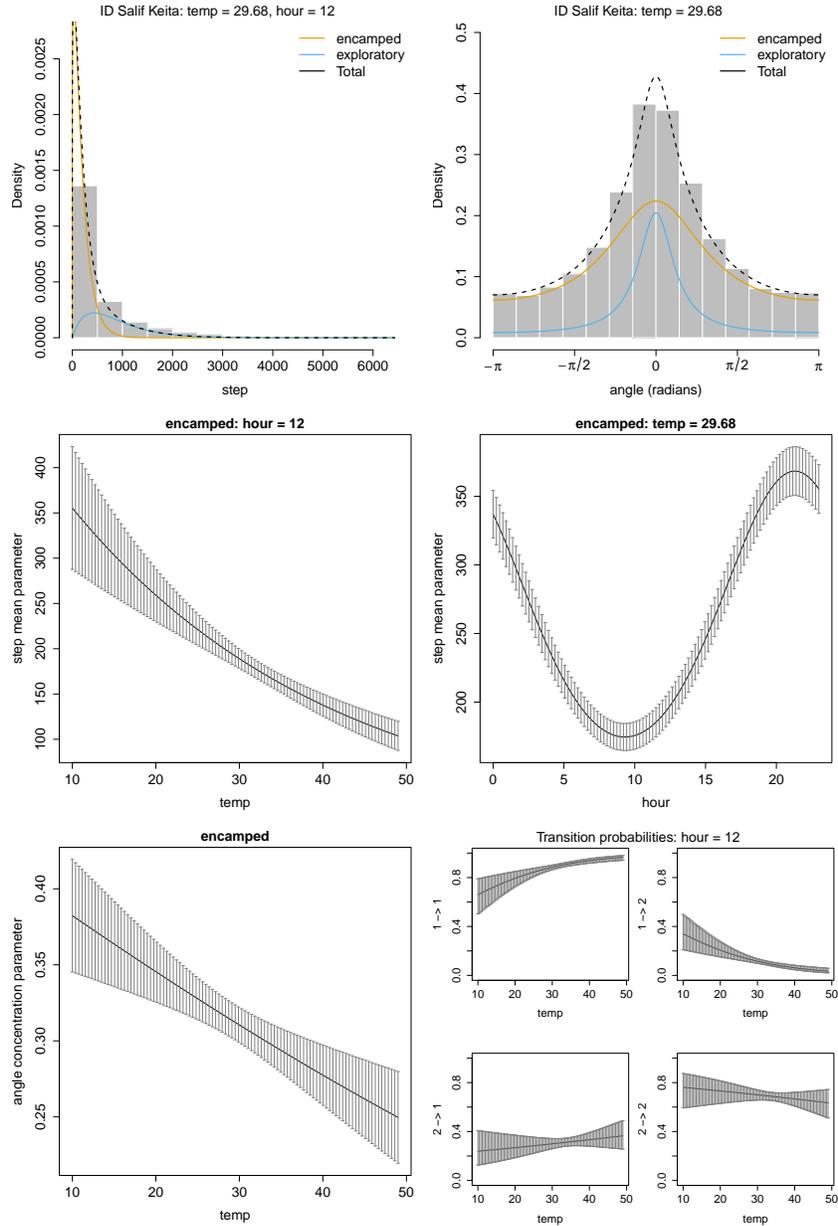
Figure 3. *Estimated probability distribution and parameter plots for the 2-state ("encamped" and "exploratory") African elephant example generated using the generic 'plot' function. Top panels present histograms of the step length (top-left) and turning angle (top-right) data along with the estimated state-dependent probability distributions based on the mean temperature (temp = 29.7 degrees celsius) at 12:00 GMT (hour = 12). Middle panels present estimates (and 95% confidence intervals) for the step length mean parameter of the "encamped" state as a function of temperature and hour of day. Bottom-left panel presents estimates for the turning angle concentration parameter of the "encamped" state as a function of temperature. Bottom-right panel presents estimated state transition probabilities (1 = "encamped", 2 = "exploratory") as a function of temperature at 12:00 GMT.*
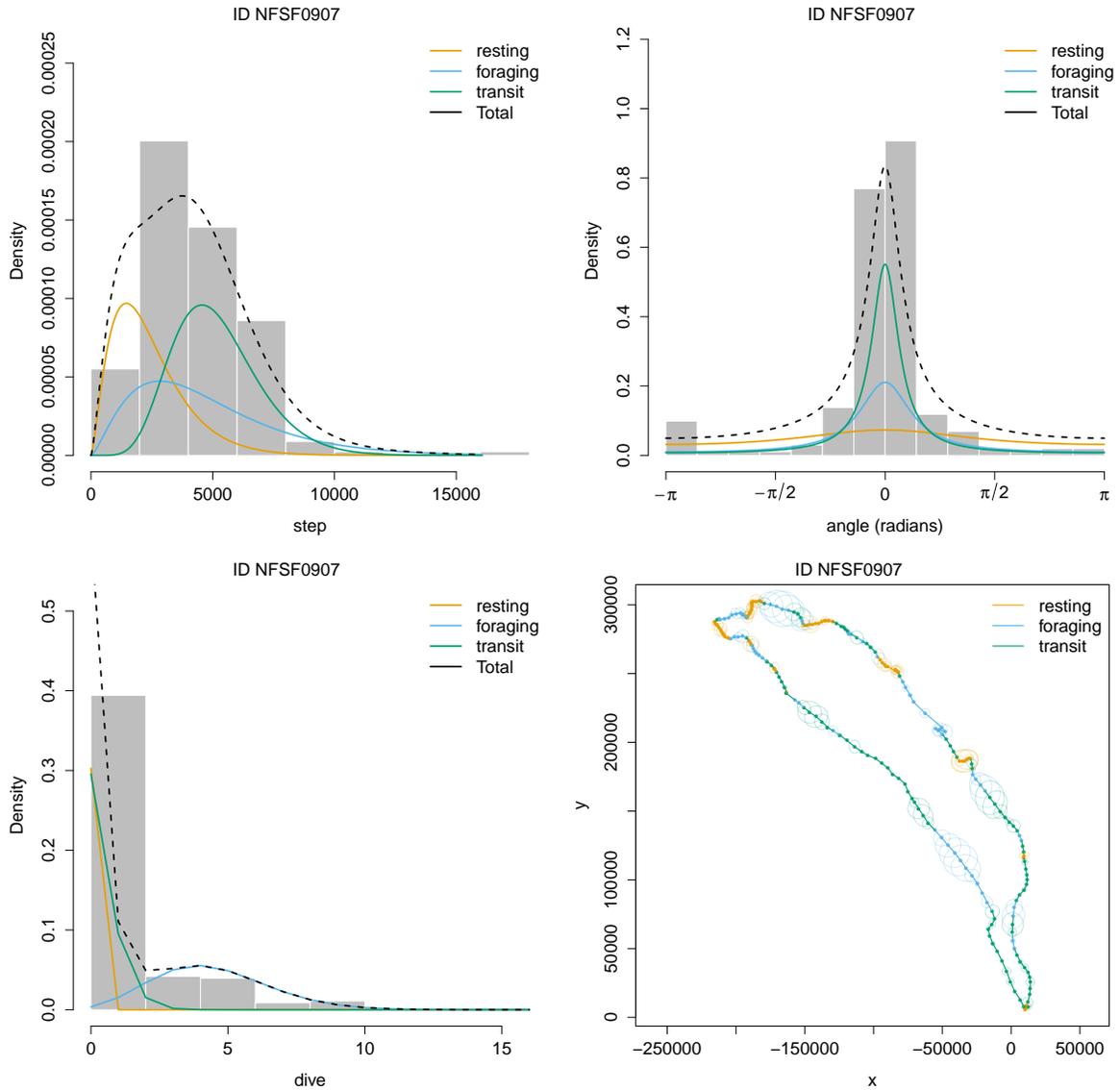
Figure 4. *Plots of the northern fur seal example results generated using the generic 'plot' function. The estimated probability distributions for step length (top-left panel), turning angle (top-right panel), and number of foraging dives (bottom-left panel) for the 3-state ("resting", "foraging", and "transit") model are plotted along with histograms of these data streams. The temporally-regular predicted locations, 95% ellipsoidal confidence bands, and estimated states are plotted in the bottom-right panel. All estimates are pooled across multiple imputations of the position process and thus reflect uncertainty attributable to location measurement error and temporally-irregular observations.*

daily ocean surface currents collected between 20 November and 19 December 2012. Assuming a gamma distribution for step length and a wrapped Cauchy distribution for turning angle, we modelled the mean step length parameter as a function of ocean surface current speed (w_t) and direction (r_t) relative to the bearing of movement (b_t) at time t. We also modelled the turning angle mean parameter as a trade-off between short-term directional persistence and bias in the direction of ocean surface currents using the circular-circular regression link function.

Using multiple imputation, we fit a 2-state HMM including a "foraging" state unaffected by currents and a "transit" state potentially influenced by ocean surface currents. After preparing the data, this rather complicated HMM can be specified, fitted, and visualized in only a few lines of code (see package vignette for complete details). For the "transit" state, pooled parameter estimates indicated step lengths increased with ocean surface current speed and as the bearing of movement aligned with ocean surface current direction (Fig. 5). The estimated wrapped Cauchy distribution for turning angle had mean angles (μ_t) biased towards the direction of ocean surface currents for each time step, with concentration parameter $\rho = 0.86$ (95% CI: $0.8-0.92$) indicating turning angles were highly concentrated around μ_t. Thus movement during the "transit" state appears to strongly follow ocean surface currents, while movement during the "foraging" state exhibited shorter step lengths perpendicular to ocean surface currents with no directional persistence.

3.4   Grey seal

We now revisit an analysis of a grey seal (*Halichoerus grypus*) track that was originally conducted by McClintock *et al.* (2012) using Bayesian methods and computationally-intensive Markov chain Monte Carlo. This seal tended to move in a clockwise fashion between two terrestrial "haul-out" sites and a foraging area in the North Sea. Because the seal repeatedly visited the same haul-out and foraging locations, it provides an excellent example for demonstrating how biased movements relative to activity centres can be modelled using `momentuHMM`. McClintock *et al.* (2012) fitted a 5-state HMM to these data that included three "centre of attraction" states, with movement biased towards two haul-out sites ("Abertay" and "Farne Islands") and a foraging area ("Dogger Bank"), and two "exploratory" states that were unassociated with an activity centre. After using `crawlWrap` to predict $T = 1515$ temporally-regular locations at 2 h time
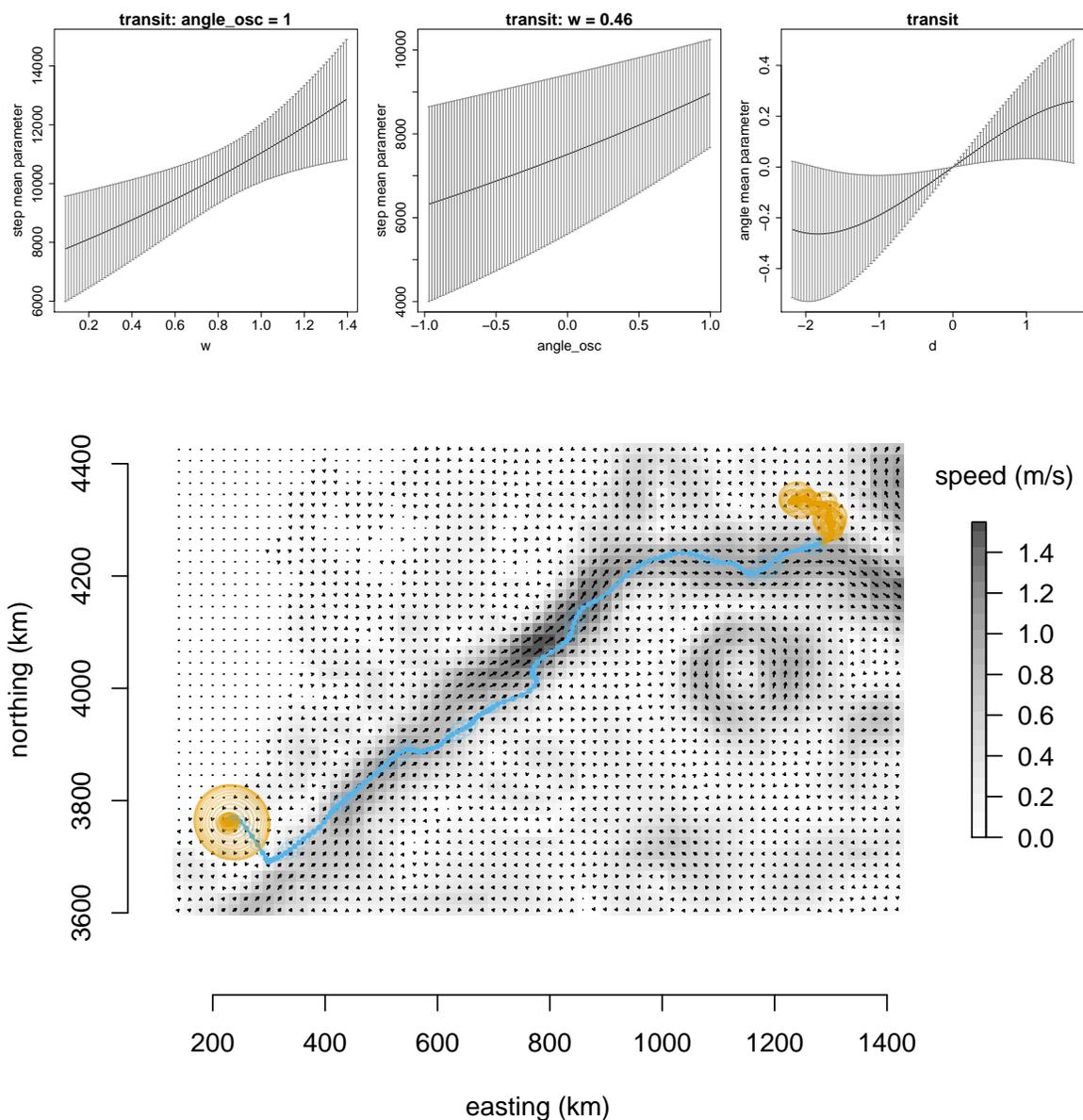
Figure 5. *Selected results from the loggerhead turtle example. Top panels include estimates and 95% confidence intervals for the "transit" state mean step length parameter as a function of ocean surface current speed (w) when ocean surface current direction (r_t) is the same as the bearing (b_t) of movement (i.e. angle_osc = $\cos(b_t - r_t) = 1$; top-left panel), mean step length parameter as a function of angle_osc at the mean ocean surface current speed ($w = 0.46$ m/s; top-middle panel), and mean turning angle parameter as a function of $d_t = \text{atan2}(\sin(r_t - b_{t-1}), \cos(r_t - b_{t-1}))$ (top-right panel). Bottom panel plots the pooled track, 95% error ellipse confidence band, and state (orange = "foraging", blue = "transit") estimates based on multiple imputations of the position process relative to ocean surface current speed (m/s) and direction on 2 December 2012. Estimates therefore reflect uncertainty attributable to location measurement error and temporally-irregular observations. The turtle spent 0.57 (95% CI: 0.48−0.65) of the 2 h time steps in the "foraging" state and 0.43 (95% CI: 0.35−0.52) of time steps in the "transit" state as it travelled northeast along a predominant current until it (presumably) found an attractive foraging patch.*

18

steps, we performed a very similar analysis to McClintock *et al.* (2012) using `MIfitHMM`.

Each model fit required about 4 min on a standard desktop computer (macOS El Capitan, 16 GB RAM, 2.8 GHz Intel Core i7). Estimated activity budgets for the 5 states of this multiple-imputation HMM were 0.28 ($0.28 - 0.29$) for the "Abertay" haul-out state, 0.12 ($0.11 - 0.14$) for the "Farne Islands" haul-out state, 0.37 ($0.35 - 0.38$) for the "Dogger Bank" foraging state, 0.09 ($0.04 - 0.18$) for a low-speed "exploratory" state, and 0.14 ($0.09 - 0.22$) for a high-speed "exploratory" state. All three activity centre states exhibited shorter step lengths and less biased movements when within the vicinity of these targets. Results from this analysis were thus very similar to McClintock *et al.* (2012), but this implementation required far less computation time and no custom model-fitting algorithms. A simulated track generated using `simData` is presented along with the fitted track in Fig. 6.

4   Discussion

We have introduced the R package `momentuHMM` and demonstrated some of its capabilities for conducting multivariate HMM analyses with animal location, auxiliary biotelemetry, and environmental data. There are many types of analyses that can be conducted with `momentuHMM` that hitherto required custom code and model-fitting algorithms (e.g. McClintock *et al.* 2013, 2017; Langrock *et al.* 2014; Michelot *et al.* 2017). The package therefore greatly expands on available software and facilitates the incorporation of more ecological and behavioural realism for hypothesis-driven analyses of animal movement that account for many of the challenges commonly associated with telemetry data. While many features were motivated by animal movement data, we note that the package can be used for analyzing any type of data that is amenable to HMMs.

Model fitting is relatively fast because the forward algorithm (Eq. 1) is coded in C++. Because multiple imputations are completely parallelizable, with sufficient processing power computation times for analyses that account for measurement error, temporal irregularity, or other forms of missing data need not be longer than that required to fit a single HMM. However, computation times will necessarily be longer as the number of states and/or parameters increase. For example, on a standard desktop computer (macOS El Capitan, 16 GB RAM, 2.8 GHz Intel Core i7), `momentuHMM` re-
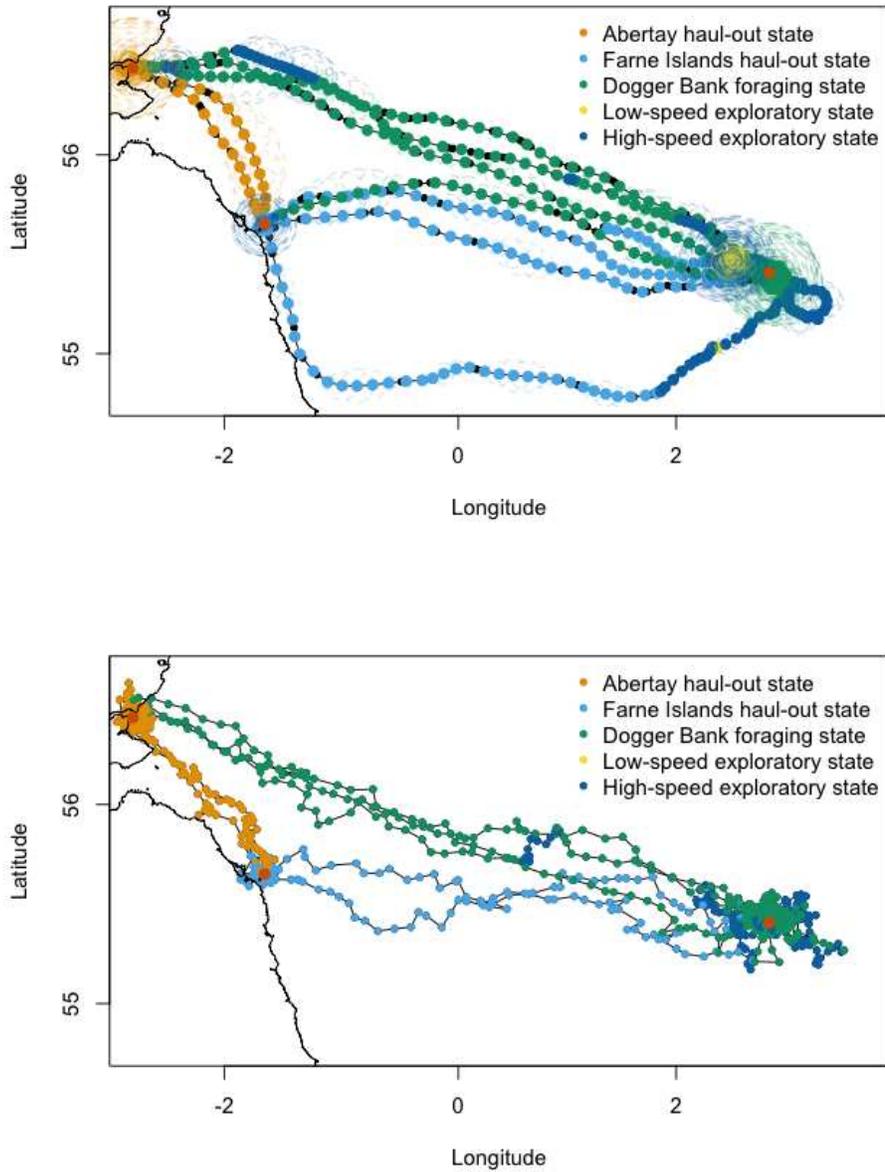
Figure 6. *Fitted and simulated tracks from the grey seal example. This seal tended to move in a clockwise fashion between two haul-out locations ("Abertay" and "Farne Islands") and a foraging area ("Dogger Bank") in the North Sea. Top panel plots the pooled track, 95% error ellipse confidence bands, and state estimates based on the 5-state HMM fitted to multiple imputations of the position process. Red points indicate the locations of the three activity centres. Black points indicate the (temporally-irregular) observed locations. Bottom panel presents the locations and states for a track simulated from the fitted model using the 'simData' function.*

quired about 1 hour to fit a single HMM with $N = 6$ states, seven data streams, and $T = 7414$ time steps (McClintock 2017).

As in any maximum likelihood analysis based on numerical optimization, computation times will also depend on starting values. Specifying "good" starting values is arguably the most challenging aspect of model fitting, and the package includes functions that are designed to help with the specification of starting values (e.g. `checkPar0`, `getPar`, `getPar0`, and `getParDM`). It also includes options for re-optimization based on random perturbations of the parameters to help explore the likelihood surface and diagnose convergence to local maxima.

While `momentuHMM` includes functions for drawing realizations of the position process based on the CTCRW model of Johnson *et al.* (2008), this is but one of many methods for performing multiple imputation. Realizations of the position process from any movement model that accounts for measurement error and/or temporal irregularity (e.g. Calabrese *et al.* 2016) can be passed to `MIfitHMM`. Multiple imputation methods also need not be limited to these telemetry error scenarios. For example, other forms of missing data could be imputed, thereby allowing investigation of non-random mechanisms for missingness that can be problematic if left unaccounted for in HMMs.

There remain many potential avenues for refinement and extension. Computation times could likely be improved by further optimizing the code for speed. Notable extensions include hidden semi-Markov models and random effects (Zucchini *et al.* 2016). It is straightforward to add additional data stream probability distributions, and practitioners interested in additional features are encouraged to contact the authors.

Acknowledgments

Authors' contributions

B.T.M. and T.M. developed the package and wrote the manuscript.

Data accessibility

All code mentioned here is available in the `momentuHMM` package for R hosted on CRAN at `https://CRAN.R-project.org/package=momentuHMM`. The development version of the package is available on GitHub at `https://github.com/bmcclintock/momentuHMM`.

References

Beyer, H.L., Morales, J.M., Murray, D. & 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.

Calabrese, J.M., Fleming, C.H. & Gurarie, E. (2016) ctmm: an R package for analyzing animal relocation data as a continuous-time stochastic process. *Methods in Ecology and Evolution*, **7**, 1124–1132.

Costa, D.P., Robinson, P.W., Arnould, J.P., Harrison, A.L., Simmons, S.E., Hassrick, J.L., Hoskins, A.J., Kirkman, S.P., Oosthuizen, H., Villegas-Amtmann, 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., Calambokidis, J., Friedlaender, A.S. & Southall, B.L. (2017) A multivariate mixed hidden Markov model to analyze blue whale diving behaviour during controlled sound exposures. *The Annals of Applied Statistics*, **11**, 362–392.

Hijmans, R.J. (2016) *raster: Geographic Data Analysis and Modeling.* R package version 2.5-8.

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

Johnson, D.S. (2017) *crawl: Fit Continuous-Time Correlated Random Walk Models to Animal Movement Data.* R package version 2.1.1.

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

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

Langrock, R., Hopcraft, G., Blackwell, P., Goodall, V., King, R., Niu, M., Patterson, T., Pedersen, M., Skarin, A. & Schick, R. (2014) Modelling group dynamic animal movement. *Methods in Ecology and Evolution*, **5**, 190–199.

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

McClintock, B.T. (2017) Incorporating telemetry error into hidden markov models of animal movement using multiple imputation. *Journal of Agricultural, Biological, and Environmental Statistics*, **22**, 249–269.

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

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

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

McCullagh, P. & Nelder, J.A. (1989) *Generalized Linear Models, Second Edition*. Chapman and Hall, New York.

Michelot, T., Langrock, R., Bestley, S., Jonsen, I.D., Photopoulou, T. & Patterson, T.A. (2017) Estimation and simulation of foraging trips in land-based marine predators. *Ecology*, **98**, 1932–1944.

Michelot, T., Langrock, R. & Patterson, T.A. (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.

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

R Core Team (2017) *R: A Language and Environment for Statistical Computing.* R Foundation for Statistical Computing, Vienna, Austria.

Rivest, L.P., Duchesne, T., Nicosia, A. & Fortin, D. (2016) A general angular regression model for the analysis of data on animal movement in ecology. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, **65**, 445–463.

Wall, J., Wittemyer, G., LeMay, V., Douglas-Hamilton, I. & Klinkenberg, B. (2014) Elliptical time-density model to estimate wildlife utilization distributions. *Methods in Ecology and Evolution*, **5**, 780–790.

Whoriskey, K., Auger-Méthé, M., Albertsen, C.M., Whoriskey, F.G., Binder, T.R., Krueger, C.C. & Mills Flemming, J. (2017) A hidden markov movement model for rapidly identifying behavioral states from animal tracks. *Ecology and Evolution*, **7**, 2112–2121.

Zucchini, W., MacDonald, I.L. & Langrock, R. (2016) *Hidden Markov Models for Time Series: An Introduction Using R.* CRC Press.