

1 **foieGras** an R package for animal movement data:
2 rapid quality control, behavioural estimation and
3 simulation

5 **Ian D. Jonsen^{1,*}, W. James Grecian², Lachlan Phillips¹, Gemma Carroll³, Clive
6 McMahon⁴, Robert G. Harcourt¹, and Toby A. Patterson⁵**

7 ¹School of Natural Sciences, Macquarie University, Sydney, NSW, Australia

8 ²Sea Mammal Research Unit, Scottish Oceans Institute, University of St Andrews, St
9 Andrews, Fife, United Kingdom

10 ³Environmental Defense Fund, Seattle, WA, United States

11 ⁴Sydney Institute of Marine Science, Mosman, NSW, Australia

12 ⁵CSIRO Ocean and Atmosphere Research, Hobart, TAS, Australia

13 *corresponding author, ian.jonsen@mq.edu.au

14 **Abstract**

- 15 1.
16 2.
17 3.
18 4.

19 **Keywords:**

20 **1 | Introduction**

21 Animal biotelemetry as a discipline has matured, with telemetry data now virtually essen-
22 tial for understanding behaviour and social interactions, foraging ecology, habitat use and
23 population dynamics of mobile and/or cryptic species. Additionally, the sophistication of cur-
24 rent telemetry devices enables the use of animal-borne sensors as a cost-effective approach
25 for observing our planet that complements more traditional observing platforms (Harcourt
26 et al., 2019; Kays et al., 2015; McMahon et al., 2021). In all these applications, animal
27 biotelemetry requires rigorous quality control procedures to account for common, though
28 not universally present, data issues such as irregularly timed measurements, sensor biases
29 and location measurement error. Some of these issues may be handled by a manufacturer's
30 on-board or subsequent processing and some must be dealt with by researchers using the
31 data.

32 State-space models (SSMs) and hidden Markov models (HMMs) are powerful tools for con-
33 ducting quality control of and making behavioural inference from animal biotelemetry data
34 (Jonsen et al., 2013; Patterson et al., 2008). These are time-series models used across a

wide range of research disciplines that estimate the state of an unobserved process from an observed data set. Here, we view an animal's true location and/or behaviour as the unobserved state(s), though many other types of states are possible (Hooten et al., 2019; e.g., Schick et al., 2013), and measurements recorded by biotelemetry devices provide the observations. In practical yet simplistic terms, SSMs are usually preferred when the goal is to quality control error-prone location data and estimate parameters of underlying movement process models. HMMs are usually preferred when the goal is to infer behavioural states hidden within biotelemetry data and their drivers, where measurements have negligible error and occur at regular time intervals (but see McClintock & Michelot, 2018). Other more technical distinctions and reasons for preferring one of these methods exist (Jonsen et al., 2013; Patterson et al., 2017). Our primary focus here is on SSMs as tools for quality control of error-prone location data.

A number of R packages such as `moveHMM` (Michelot et al., 2016), `momentuHMM` (McClintock & Michelot, 2018), and `swim` (Whoriskey et al., 2017) provide highly accessible and flexible tools for fitting HMMs to biotelemetry data, and facilitating general inference of animal movement behaviour and its drivers. Similarly, R packages such as `bsam` (I. Jonsen et al., 2005), `crawl` (Johnson et al., 2008), `argosTrack` (Albertsen et al., 2015), `ctmm` (Calabrese et al., 2016), and `yaps` (Baktoft et al., 2017) all provide tools for fitting movement process models in either discrete- or continuous-time, ranging from simple random walks to Ornstein-Uhlenbeck processes, in state-space form to various types of animal biotelemetry data.

The `foieGras`, pronounced “*fwah grah*”, package for R (R Core Team, 2021) was developed to be simple and fast to use, providing SSMs for quality control of error-prone (re)location data collected via the Argos satellite [Service Argos (2016); Jonsen:2020] and other telemetry systems, and for inference of changes in behaviour along movement tracks (Jonsen et al., 2019). The simplicity of use sets `foieGras` apart from many of the related SSM R packages listed above, yet users can exert control over many aspects of the package functions via optional arguments. This design accommodates both novice and experienced users, and facilitates use in automated, operational quality control workflows (Jonsen et al., 2020).

Here, we describe the main features of `foieGras` and illustrate its use through a set of examples using Argos and GPS data. Full R code and data for each of the examples is provided in the Supporting Information. Additional details on package functions and their use can be found in their help files and in the package’s vignettes.

2 | `foieGras` overview

The workflow for `foieGras` is deliberately simple, with many of the usual track data processing checks and formatting handled automatically. The main functions are listed in Table 1. When fitting a model, `foieGras` automatically detects the type of tracking data from the location quality class designations that are typical of Argos data and that can be added to the data by the researcher for other types of track data. Based on the location quality classes and other, optional information on measurement errors contained in the data, `foieGras` chooses an appropriate measurement error model for each observation (Jonsen et al., 2020). This capability can allow different tracking data types, e.g., Argos and GPS, to be combined in a

Table 1: Main functions for the R package **foieGras**

Function	Description
<code>fit_ssm</code>	Fit a State-Space Model to location data
<code>fit_mpm</code>	Fit a Move Persistence Model to location data
<code>grab</code>	Extract fitted/predicted/observed locations from a foieGras model, with or without projection information
<code>osar</code>	Estimate One-Step-Ahead Residuals from a foieGras SSM
<code>map</code>	Map fitted/predicted locations with or without a defined projection
<code>sim</code>	Simulate individual animal tracks with Argos LS or KF errors
<code>simfit</code>	Simulate animal tracks from ‘fG_ssm’ fit objects
<code>sim_filter</code>	Filter tracks simulated with ‘simfit’ according to similarity criteria
<code>route_path</code>	Reroute path so estimated locations are off land
<code>plot.ssm_df</code>	Plot the fit of a foieGras SSM to data
<code>plot.osar</code>	Plot One-Step-Ahead Residuals from a foieGras SSM
<code>plot.mpm_df</code>	Plot move persistence estimates as 1-D or 2-D (along track) time-series
<code>plot.sim</code>	Plot simulated animal tracks

⁷⁶ single input data frame and to be fit in a single state-space model.

⁷⁷ Estimation is performed via maximum likelihood using the **TMB** R package (Kristensen et al.,
⁷⁸ 2016) and R’s standard optimizers `optim` or `nlminb`. **TMB** allows the SSMs to be specified
⁷⁹ as C++ templates to rapidly evaluate the model’s negative log-likelihood function using
⁸⁰ reverse-mode auto-differentiation and the Laplace approximation for handling random effects
⁸¹ (Kristensen et al., 2016).

⁸² 2.1 | Data preparation

⁸³ Animal tracking data, consisting of a time-series of location coordinates, can be read into R
⁸⁴ as a data frame using standard functions such as `read.csv`. The canonical data format for
⁸⁵ Argos tracks consists of a data frame with 5 columns corresponding to the following named
⁸⁶ variables: `id` (individual id), `date` (date and time), `lc` (location class), `lon` (longitude), `lat`
⁸⁷ (latitude). Optionally, an additional 3 columns, `smaj` (semi-major axis), `smin` (semi-minor
⁸⁸ axis), `eor` (ellipse orientation), providing Argos error ellipse information may be included.

⁸⁹ Other types of track data can be accommodated, for example, by including the `lc` column
⁹⁰ where all `lc = "G"` for GPS data. In this case, measurement error in the GPS locations is
⁹¹ assumed to have a standard deviation of 0.1 x Argos class 3 locations (approximately 30 m).
⁹² Other types of track data may be also be considered, including those derived from light-level

93 geolocation (see the package vignette for further details).

94 **2.2 | State-space model fitting - `fit_ssm`**

95 State-space models are fit using `fit_ssm`. There are a large number of options that can be
96 set in `fit_ssm` (see Suppl for details). We focus only the essential options here:

- 97 • `data` the input data structured as described in **2.1**
- 98 • `vmax` a maximum threshold speed (ms^{-1}) to help identify potential outlier locations
- 99 • `model` the process model to be used
- 100 • `time.step` the prediction time interval (h)

101 The function first invokes an automated data processing stage where the following occurs:
102 1) data type (Argos Least-Squares, Argos Kalman Filter/Smoother, GPS, or General (e.g.,
103 processed light-level geolocations, acoustic telemetry, coded VHF telemetry) is determined;
104 2) datetimes are converted to POSIXt format, chronological order is ensured, and duplicate
105 datetime records are removed; 3) observations occurring less than `min.dt` seconds after a
106 prior observation are removed; 4) a speed filter [`sda` from the `trip` R package; Sumner et al.
107 (2009)] is used to identify potential outlier locations; 5) locations are projected from spherical
108 lon-lat coordinates to planar x,y coordinates in km.

109 The function then fits a state-space model to the processed data, where the process model
110 (currently, either a continuous-time `rw` or a continuous-time `crw`) is specified by the user
111 and the measurement model(s) are selected automatically (see Jonsen et al., 2020 for model
112 details). The model is fit by numerical optimization of the likelihood using either the `optim`
113 or `nlminb` R function. The R package `TMB`, Template Model Builder (Kristensen et al.,
114 2016), is used to compute the gradient function in C++ via reverse-mode auto-differentiation
115 and the Laplace Approximation is used to integrate out the latent states (random effects).
116 Fits to a single versus multiple individuals are handled automatically, with sequential SSM
117 fits occurring in the latter case. No hierarchical or pooled estimation among individuals is
118 currently available.

119 `fit_ssm` returns a `foieGras` fit object (a nested data frame with class `fG_ssm`). The outer
120 data frame lists the individual id(s), basic convergence information and a list with class `ssm`.
121 This list contains dense information on the model parameter and state estimates, predictions,
122 processed data, optimizer results, and other diagnostic and contextual information. Users can
123 extract a simple data frame of SSM fitted (location estimates corresponding to the, typically
124 irregular, observation times) or predicted values (locations predicted at regular `time.step`
125 intervals) using the `grab` function.

126 **2.3 | Model checking and visualisation - `osar`, `plot`, `fmap`**

127 Before using fitted or predicted locations, a model fit should be checked and visualised to
128 confirm that the model adequately describes the data. In linear regression and a variety of
129 analogous methods, goodness-of-fit can be assessed by calculating standard residuals such
130 as Pearson or deviance residuals. There is no simple way to calculate residuals for latent
131 variable models that have non-finite state-spaces and that may be nonlinear, but they can be

132 computed based on iterative forecasts of the model (Thygesen et al., 2017). The `osar` function
133 computes one-step-ahead (prediction) residuals and uses the `oneStepPredict` function from
134 the `TMB` R package to make this as efficient as possible. A set of residuals are calculated
135 for the `x` and `y` values corresponding to the fitted values from the SSM and returned as an
136 `fG_osar` object.

137 A generic `plot` method provides an easy way to visualise the `fG_osar` residuals. Time-series
138 plots of the prediction residuals can be used to detect temporal changes in goodness-of-fit.
139 Quantile-quantile plots of residuals against standard normal quantiles can be used to detect
140 departures from normality. Sample autocorrelation function plots of the residuals are useful
141 for detecting autocorrelation not accounted for by the model. Assessing residual autocorre-
142 lation can be particularly important as Argos locations, for example, are themselves derived
143 from a time-series model (Lopez et al., 2015) which can introduce additional autocorrelation
144 in the location errors.

145 State-space model fits to data can also be visualised by using the generic `plot` function on an
146 `fG_ssm` data frame. Options exist to plot fitted or predicted values along with observations as
147 either paired, 1-D time-series or as 2-D tracks with confidence intervals or ellipses, respectively.
148 These plots provide a more intuitive and rapid method for assessing SSM fits to data, however,
149 they do not replace the residual diagnostics. Fitted `fG_ssm` data frames can be mapped using
150 the `fmap` function for single or multiple individuals. Estimated tracks can be displayed with
151 or without confidence ellipses, observations, and/or a projection and maps of single tracks
152 can be coloured by date.

153 2.4 | Behavioural estimation - `fit_mpm`

154 The `fit_mpm` function fits a simple move persistence model to estimate a continuous-valued,
155 time-varying latent variable that indexes changes in movement behaviour (Jonsen et al., 2019).
156 This variable measures the autocorrelation in speed and direction between consecutive pairs
157 of movements such that high values correspond to fast, directed movements at one end of
158 the continuum and low values correspond to slow, tortuous movements at the other end. It's
159 important to note that this approach is unlike hidden Markov models (McClintock & Michelot,
160 2018; Michelot et al., 2016) and some state-space models (Jonsen, 2016) as there is no notion
161 of discrete behavioural states that animals periodically switch between. Nonetheless, move
162 persistence can be used to identify objectively places where animals spend disproportionately
163 more or less time, and with extensions be correlated with environment or other covariates
164 (See Examples 3.x).

165 The move persistence model assumes that locations are absent of measurement error and can
166 occur either irregularly or regularly in time. `fit_mpm` takes either a `fG_ssm` data frame as
167 input or a data frame with the follow variables: `id`, `date`, `x`, `y`, where `x` and `y` coordinates
168 can be planar `x,y` or spherical `long,lat`. This latter input format allows the model to be
169 fit easily to GPS or other tracking data with negligible measurement error. When the data
170 contain multiple individuals, the default model is fit jointly by assuming all individuals share
171 the same move persistence variance parameter (Jonsen, 2016). There is an option to fit the
172 model separately to each individual. The time-series of estimated move persistence with
173 confidence intervals can be visualized by using the generic `plot` function with the resulting

174 `fG_mpm` data frame. Visualization of move persistence along the 2-D tracks can be plotted or
175 mapped by using the `plot` or `fmap` functions, respectively, and supplying both the `fG_mpm` and
176 `fG_ssm` nested data frames. When using `fit_mpm` on, for example, GPS tracking data that do
177 not require state-space filtering, the movement persistence estimates can be extracted from
178 the `fG_mpm` data frame using the `grab` function and subsequently merged with the observed
179 track data for visualization.

180 **2.5 | Simulation - `sim`, `simfit`, `sim_filter`**

181 Track simulation can be a helpful, yet informal, way of evaluating the degree to which statis-
182 tical movement models capture essential features of animal movement data (Michelot et al.,
183 2017). Michelot et al. (2016) advocate comparison of simulated tracks from fitted hidden
184 Markov models to the observed tracks as a means of identifying potential weakness in the
185 hidden Markov model formulation. We suggest that the `rw` and `crw` state-space models and
186 the `mpm` model can be fit to track data simulated from different movement processes to eval-
187 uate robustness of location and movement persistence estimates to model mis-specification.
188 We illustrate this idea in a Supplement (xx) by drawing on flexibility in the `sim` function
189 that allows a variety of movement processes to be simulated.

190 Simulation is also used frequently to provide a measure of habitat availability (Aarts et
191 al., 2012) by providing a source of ‘background’ points representing a null model of the
192 distribution of foraging animals in the absence of external drivers (Hindell et al., 2020; S.
193 J. Phillips et al., 2009; Raymond et al., 2015). The `simfit` function extracts movement
194 parameters from a `fG_ssm` fit object and simulates a user defined number of random tracks
195 of the same duration from these parameters. The argument `cpf = TRUE` allows the user
196 to simulate a central place forager by ensuring that the simulated tracks start and end at
197 approximately the same location.

198 The choice of null points can have a large impact on the performance of habitat suitability
199 models (Lobo et al., 2010; S. J. Phillips et al., 2009), and so the `sim_filter` function provides
200 a tool to filter the simulated tracks based on their similarity to the original path. The filtering
201 is based on one of two metrics that capture the difference in the net displacement and bearing
202 between the two paths (see `similarity_flag` for more detail). This metric is motivated by
203 the ‘flag value’ described in Hazen et al. (2017). The user can also specify the quantile of flag
204 values to retain; i.e. `keep = 0.25` (the default) will return a `simfit` object containing those
205 simulated tracks with flag values in the top 25% of values calculated for the input `simfit`
206 object.

207 **2.6 | Path rerouting - `route_path`**

208 **3 | Examples**

209 We illustrate the main capabilities of `foieGras` through a series of examples that are for
210 demonstration purposes and not intended as a comprehensive guide for conducting analyses
211 with `foieGras`. Complete code and data for reproducing the examples and for gaining a
212 deeper understanding of `foieGras` functions are provided as supplements.

213 **3.1 | Southern Elephant seal - SSM validation with prediction residuals**

214 We use a sub-adult male southern elephant seal track included in `foieGras` (`sese2`, id: ct36-
 215 E-09), sourced from from the Australian Integrated Marine Observing System (IMOS; data
 216 publicly available via imos.aodn.org.au) deployments at Iles Kerguelen in collaboration with
 217 the French IPEV and SNO-MEMO programmes. The data are temporally irregular Argos
 218 Least-Squares based locations, 74 % of which are in the poorest location quality classes: A
 219 and B. We fit both the `rw` and `crw` models using `fit_ssm` with a speed filter threshold (`vmax`)
 220 of 4 ms^{-1} and a 12-h time step. We calculate prediction residuals using `osar`, and then use
 221 the generic `plot` method for `osar` residuals to assess and compare the model fits (Fig. 1).

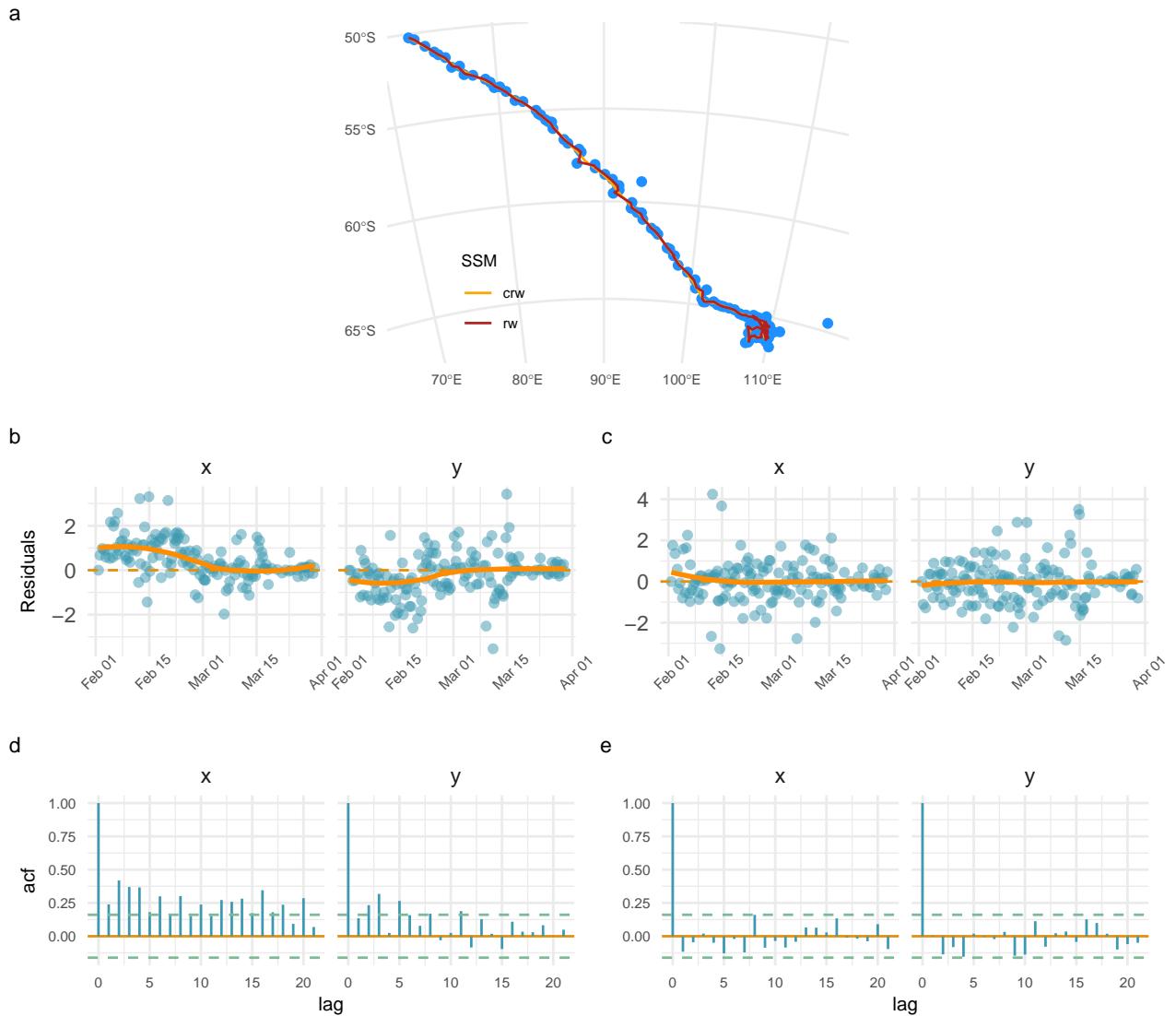


Figure 1: State-space model fits to a southern elephant seal track (a), and diagnostic plots for assessing `rw` (c - time-series of prediction residuals; e - autocorrelation of prediction residuals) and `crw` (d,f) state-space model goodness-of-fit. All residual plots generated using the `plot.fG_osar` function.

222 The plots of predicted states on top of the observations suggest both models yield similar
223 fits (Fig. 1a; orange vs red lines). However, corresponding predicted locations from the
224 two models differ by a median 6.62 km (range: 0.02, 53.02 km), and there are marked
225 trends in the time-series of residuals for the `rw` model fit (Fig. 1b) with significantly positive
226 autocorrelation in both the x and y directions (Fig. 1d). The `crw` prediction residuals show
227 little trend through time and have relatively little autocorrelation (Fig. 1c,e), implying that
228 the `crw` process model provides a better fit to the data.

229 **3.2 | Inferring movement persistence from Argos and GPS data**

230 Drawing on an expanded version of the data used in 3.1, we quality control and infer move-
231 ment persistence, γ_t , along four southern elephant seal tracks. We fitted the `mp` SSM with
232 a 24-h prediction interval using `fit_ssm`, assuming bivariate normal measurement errors
233 consistent with Argos Least-Squares-derived locations (Jonsen et al., 2020). The `mp` SSM
234 simultaneously estimates locations and γ_t , and their uncertainties. We then assess how γ_t
235 changes along the seals' tracks to infer regions where the seals spend disproportionately more
236 or less time during their foraging trips.

237 To illustrate how the method can accommodate other types of animal tracking data, we also
238 infer γ_t along four little penguin (*Eudyptula minor*) GPS tracks from Montague Island, NSW
239 Australia, described in L. Phillips et al. (2021). The data are temporally irregular GPS
240 locations, with high frequency sampling (1 - 2s) periodically disrupted by the birds' diving
241 behaviour, and are assumed to have minimal measurement error. We consequently chose to
242 fit the `crw` SSM to the GPS data and predict temporally regular locations at 5-min intervals,
243 assuming consistently small bivariate normal location measurement errors (ie. ± 10 m sd).
244 We then used `fit_mpm` to estimate γ_t from the SSM-predicted locations.

245 Movement persistence estimates along the quality-controlled southern elephant seal tracks
246 highlight some fundamental differences in movement pattern among the seals. The two seals
247 engaging in pelagic foraging trips (Fig. 2a,c and e) had less contrast in their movements
248 with consistently higher γ_t estimates compared to the two seals engaging in trips to the
249 fast-ice on the Antarctic shelf (Fig. 2b,d and e). Although γ_t 's were higher overall for the
250 pelagically foraging seals, they both spent little time making fast, highly directed movements
251 ($\gamma_t \rightarrow 1$) relative to the shelf-foraging seals (2a,c vs b,d). This suggests the pelagically-
252 foraging seals may spend considerable time searching for suitable foraging habitat in the
253 highly variable eddy fields between the Subantarctic and Polar Fronts (Jonsen et al., 2019),
254 whereas foraging habitat may be more predictable for seals travelling rapidly and directly to
255 the Antarctic shelf region. These seals may also haulout periodically on available fast-ice to
256 rest. This behaviour could also contribute to the higher contrast in movement persistence,
257 relative to pelagically-foraging seals who would not have access to fast-ice.

258 Despite vastly different scales of movement, the time series of little penguin movement persis-
259 tence estimates were broadly similar to those of the southern elephant seals (Fig. 3a-d). The
260 little penguin foraging trips likely reflect the underlying spatial distribution of their forage-
261 fish prey, with spatially diffuse bouts of lower movement persistence potentially indicative
262 of foraging both within and among neighbouring discrete prey patches (Carroll et al., 2017)
263 (Fig. 3e).

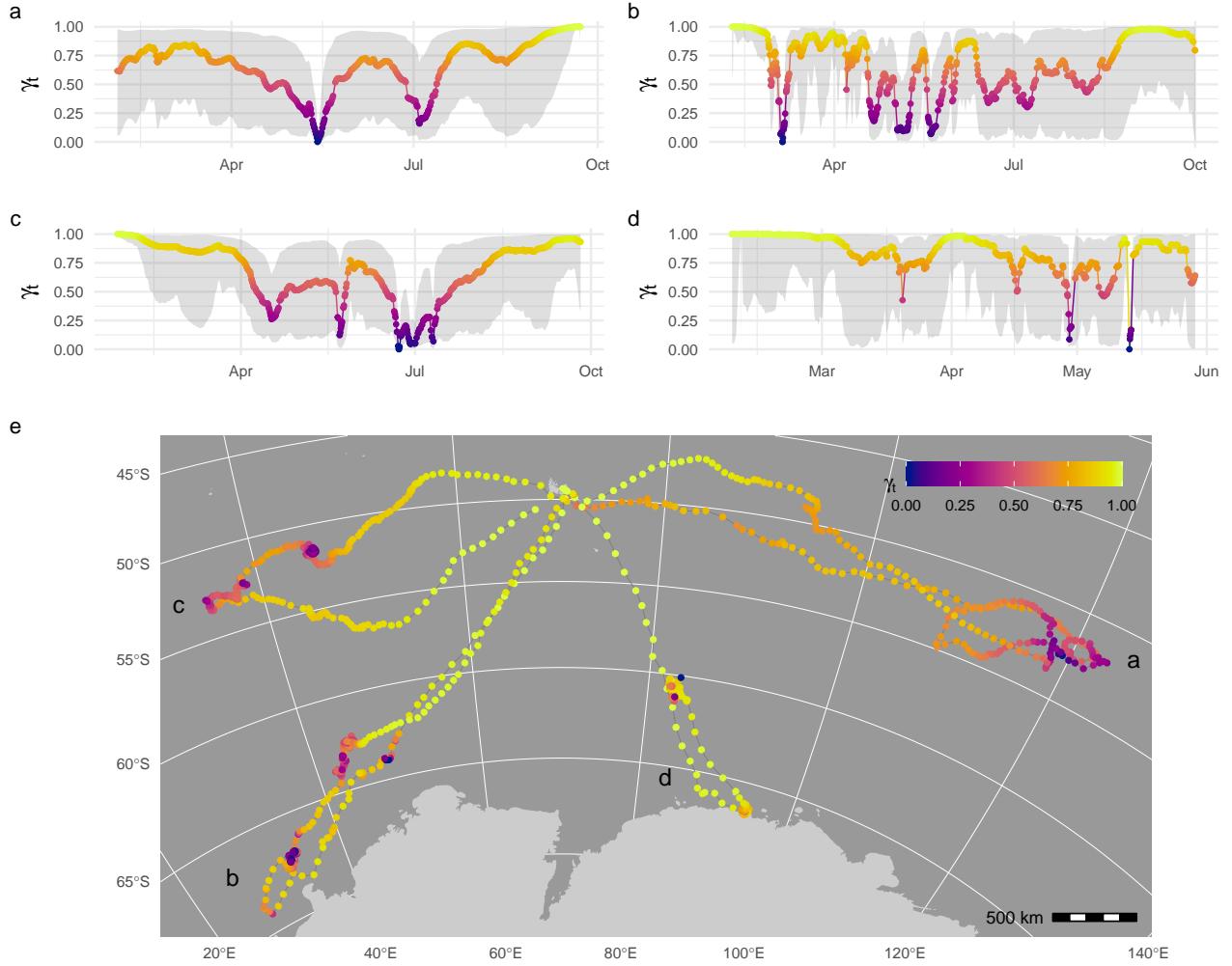


Figure 2: Inferred move persistence, γ_t , 1-D time-series for four southern elephant seals (a-d; grey envelopes are 95 % CI's) and along their 2-D tracks (e; track labels, a-d, correspond to the 1-D time-series plots). Locations associated with low move persistence (purple) are indicative of slow, undirected movements, whereas high move persistence (yellow) is indicative of faster, directed movements.

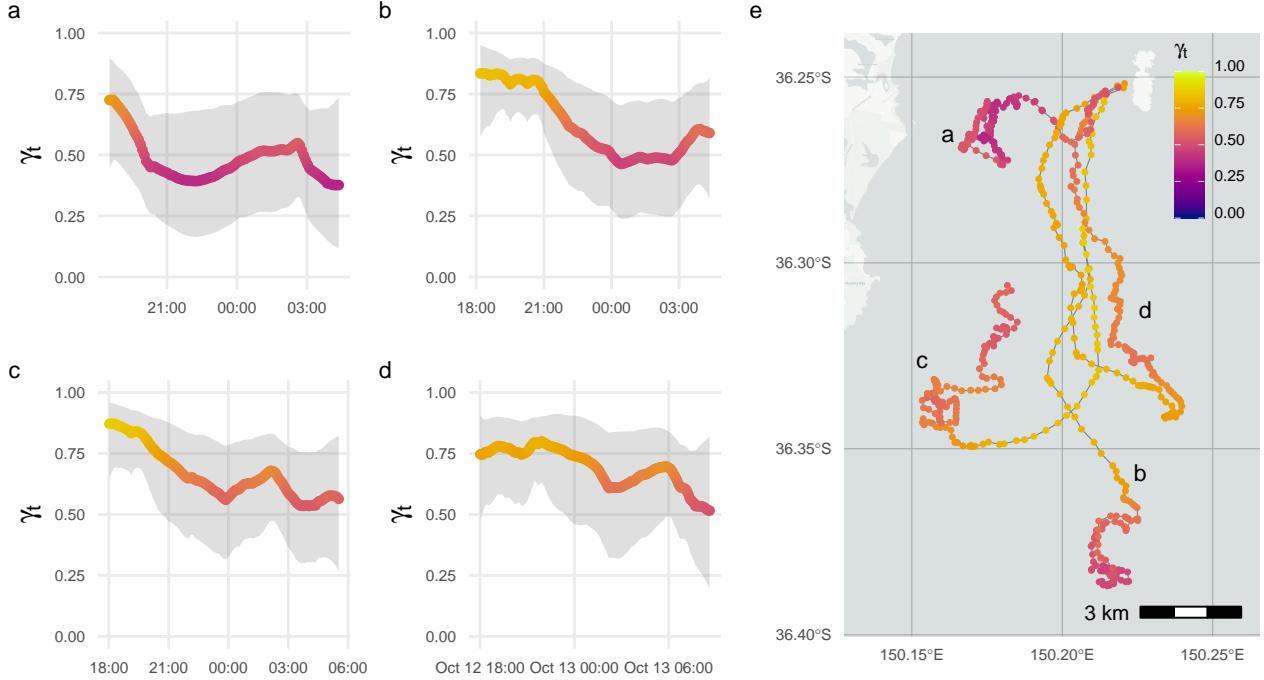


Figure 3: Inferred move persistence, γ_t , 1-D time-series (a-d; grey envelopes are 95 % CI's) and along little penguin GPS tracks (e).

264 3.3 | Simulating tracks from `foieGras` model fits

265 To illustrate how to simulate tracks from `foieGras` model fits we use a juvenile harp seal
 266 (*Pagophilus groenlandicus*) tracked from the Gulf of St Lawrence, Canada, and described in
 267 Grecian et al. (2022). The data are temporally irregular Argos locations including error
 268 ellipse information. We fit the `crw` model using `fit_ssm` with a 4 ms^{-1} speed filter threshold
 269 (`vmax`) and a 12-h prediction interval (`time.step`).

270 We simulate 50 animal movement paths from the `crw` process model using `simfit`, and apply
 271 a potential function using the `grad` and `beta` arguments to constrain the simulated paths
 272 to largely remain in water. These tracks are then filtered based on their similarity to the
 273 original path using `sim_filter` and the top 10% retained (`keep = 0.1`) (Fig. 4a,b).

274 As the potential function does not guarantee all locations remain off land, we re-route any
 275 remaining simulated locations from land back to water using `route_path` (Fig. 4c).

276 In combination, these functions provide a simple, pragmatic method to generate and objec-
 277 tively filter pseudo-tracks for use in movement or habitat modelling applications.

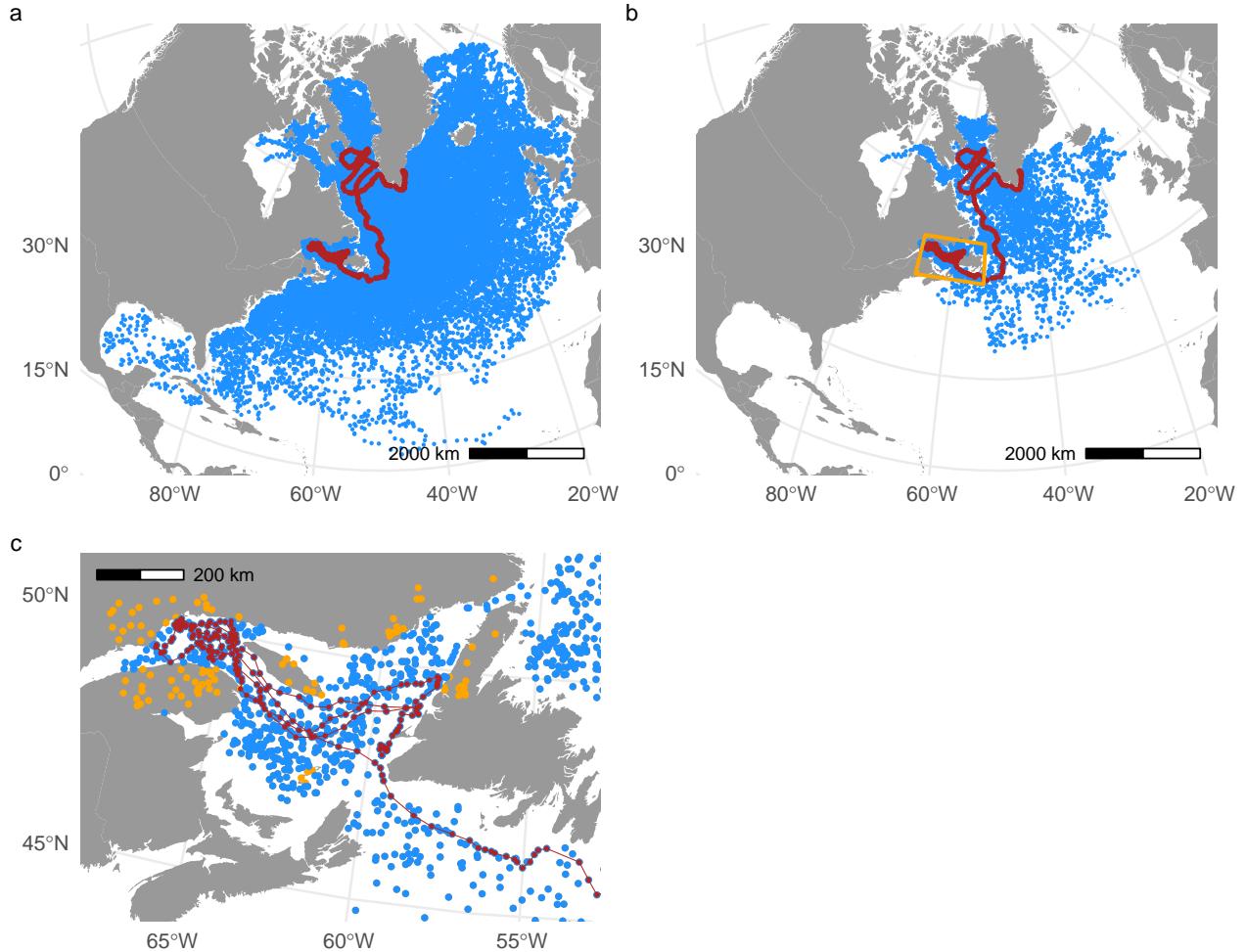


Figure 4: Simulating (a) 100 movement paths from a correlated random walk process model; (b) filtering those tracks to select the top 10% based on their similarity to the original SSM-predicted track (red); and (c) re-routing simulated locations on land (orange) back to ocean (blue). The orange box in (b) indicates region magnified in (c). SSM-predicted track (red) overlaid in all panels for context.

278 **4 | Discussion**

279 The estimation speed and simplicity of the SSMs allows them to be fit as part of a fully
280 automated, real-time quality control/quality assurances process (Jonsen et al., 2020).
281 Ex 3.2 In a limited way, this provides information on the robustness of the **foieGras** SSM's
282 to different kinds plausible animal movements

283 **Acknowledgements**

284 We thank Marie Auger-Méthé for contributing original code to the movement persistence
285 models. IDJ acknowledges support from a Macquarie University co-Funded Fellowship and
286 from partners: the US Office of Naval Research, Marine Mammal Program (grant N00014-
287 18-1-2405); the Integrated Marine Observing System (IMOS); Taronga Conservation Society;
288 the Ocean Tracking Network; Birds Canada; and Innovasea/VEMCO. TAP was supported
289 by CSIRO Oceans & Atmosphere internal research funding scheme. The Integrated Ma-
290 rine Observing System (IMOS) supported seal fieldwork. IMOS is a national collaborative
291 research infrastructure, supported by the Australian Government and operated by a consor-
292 tium of institutions as an unincorporated joint venture, with the University of Tasmania as
293 Lead Agent. Field work at Illes Kerguelen was conducted as part of the IPEV programme
294 N° 109 (PI H. WEIMERSKIRCH) and of the SNO-MEMO programme (PI C. GUINET) in
295 collaboration with IMOS. CTD tags were partly funded by CNES-TOSCA and IMOS. Little
296 penguin fieldwork was supported by an Australian Research Council Linkage grant to IDJ,
297 GC and RGH (LP160100162). All animal tagging procedures approved and executed un-
298 der the Animal Ethics Committee guidelines of the University of Tasmania (elephant seals),
299 Macquarie University (little penguins), and ... University (harp seals).

300 **Author's Contributions**

301 IDJ developed the R package; WJG contributed harp seal data and to the R package; LP,
302 GC, and RGH contributed little penguin data; CRM and RGH contributed Southern elephant
303 seal data; IDJ and TAP developed the state-space models; IDJ wrote an initial draft of the
304 manuscript with contributions from WJG; all authors edited the manuscript.

305 **Data Accessibility**

306 All code and data used here are provided in the **foieGras** package for R or in the Supple-
307 mentary Information. The latest stable and cross-platform tested version of the package
308 (currently, 1.0-6) is available via ROpenSci's R-universe, at <https://ianjonsen.r-universe.dev>
309 /ui#package:foieGras. The latest partially tested stable and development versions are avail-
310 able on the GitHub repository: <https://github.com/ianjonsen/foieGras>. An older version
311 of **foieGras** (0.7-6) remains on CRAN at <https://CRAN.R-project.org/package=foieGras>,
312 however, we recommend users upgrade to the latest R-universe version for full access to the
313 functionality presented here.

314 **ORCID**

- 315 Ian D Jonsen <https://orcid.org/0000-0001-5423-6076>
316 W James Grecian <https://orcid.org/0000-0002-6428-719X>
317 Lachlan Phillips <https://orcid.org/0000-0002-7635-2817>
318 Gemma Carroll <https://orcid.org/0000-0001-7776-0946>
319 Robert G Harcourt <https://orcid.org/0000-0003-4666-2934>
320 Clive R McMahon <https://orcid.org/0000-0001-5241-8917>
321 Toby A Patterson <https://orcid.org/0000-0002-7150-9205>

322 **References**

- 323 Aarts, G., Fieberg, J., & Matthiopoulos, J. (2012). Comparative interpretation of count,
324 presence-absence and point methods for species distribution models. *Methods in Ecology
325 and Evolution*, 3(1), 177–187. <https://doi.org/10.1111/j.2041-210x.2011.00141.x>
326 Albertsen, C. M., Whoriskey, K., Yurkowski, D., Nielsen, A., & Mills Flemming, J. (2015).
327 Fast fitting of non-Gaussian state-space models to animal movement data via Template
328 Model Builder. *Ecology*, 96, 2598–2604.
329 Baktoft, H., Gjelland, K. Ø., Økland, F., & Thygesen, U. H. (2017). Positioning of aquatic
330 animals based on time-of-arrival and random walk models using YAPS (yet another posi-
331 tioning solver). *Scientific Reports*, 7, 14294. <https://doi.org/10.1038/s41598-017-14278-z>
332 Calabrese, J. M., Fleming, C. H., & Gurarie, E. (2016). Ctmm: An R package for analyzing
333 animal relocation data as a continuous-time stochastic process. *Methods in Ecology and
334 Evolution*, 7, 1124–1132.
335 Carroll, G., Cox, M., Harcourt, R., Pitcher, B., Slip, D., & Jonsen, I. (2017). Hierarchical in-
336 fluences of prey distribution on patterns of prey capture by a marine predator. *Functional
337 Ecology*, 31, 1750–1760.
338 Grecian, W. J., Stenson, G. B., Biuw, M., Boehme, L., Folkow, L. P., Goulet, P. J., Jonsen,
339 I. D., Malde, A., Nordøy, E. S., Rosing-Asvid, A., & Smout, S. (2022). Environmental
340 drivers of population-level variation in the migratory and diving ontogeny of an arctic
341 top predator. *Royal Society Open Science*, 9(3). <https://doi.org/10.1098/rsos.211042>
342 Harcourt, R., Sequeira, A. M. M., Zhang, X., Roquet, F., Komatsu, K., Heupel, M., McMa-
343 hon, C., Whoriskey, F., Meekan, M., Carroll, G., Brodie, S., Simpfendorfer, C., Hindell,
344 M., Jonsen, I., Costa, D. P., Block, B., Muelbert, M., Woodward, B., Weise, M., ... Fedak,
345 M. A. (2019). Animal-borne telemetry: An integral component of the ocean observing
346 toolkit. *Frontiers in Marine Science*, 6, 326.
347 Hazen, E. L., Palacios, D. M., Forney, K. A., Howell, E. A., Becker, E., Hoover, A. L., Irvine,
348 L., DeAngelis, M., Bograd, S. J., Mate, B. R., & Bailey, H. (2017). WhaleWatch: a
349 dynamic management tool for predicting blue whale density in the California Current.
350 *Journal of Applied Ecology*, 54(5), 1415–1428. <https://doi.org/10.1111/1365-2664.12820>
351 Hindell, M. A., Reisinger, R. R., Ropert-Coudert, Y., Hückstädt, L. A., Trathan, P. N.,
352 Bornemann, H., Charrassin, J.-B., Chown, S. L., Costa, D. P., Danis, B.others. (2020).
353 Tracking of marine predators to protect southern ocean ecosystems. *Nature*, 580(7801),
354 87–92.
355 Hooten, M. B., Scharf, H. R., & Morales, J. M. (2019). Running on empty: Recharge

- 356 dynamics from animal movement data. *Ecology Letters*, 22, 377–389.
- 357 Johnson, D. S., London, J. M., Lea, M.-A., & Durban, J. W. (2008). Continuous-time
358 correlated random walk model for animal telemetry data. *Ecology*, 89(5), 1208–1215.
- 359 Jonsen, I. D. (2016). Joint estimation over multiple individuals improves behavioural state
360 inference from animal movement data. *Scientific Reports*, 6, 20625.
- 361 Jonsen, I. D., Basson, M., Bestley, S., Bravington, M. V., Patterson, T. A., Pedersen, M.
362 W., Thomson, R., Thygesen, U. H., & Wotherspoon, S. J. (2013). State-space models for
363 bio-loggers: A methodological road map. *Deep Sea Research II*, 88-89, 34–46.
- 364 Jonsen, I. D., McMahon, C. R., Patterson, T. A., Auger-Méthé, M., Harcourt, R., Hindell, M.
365 A., & Bestley, S. (2019). Movement responses to environment: Fast inference of variation
366 among southern elephant seals with a mixed effects model. *Ecology*, 100, e02566.
- 367 Jonsen, I. D., Patterson, T. A., Costa, D. P., Doherty, P. D., Godley, B. J., Grecian, W. J.,
368 Guinet, C., Hoennner, X., Kienle, S. S., Robinson, P. W., Votier, S. C., Whiting, S., Witt,
369 M. J., Hindell, M. A., Harcourt, R. G., & McMahon, C. R. (2020). A continuous-time
370 state-space model for rapid quality control of Argos locations from animal-borne tags.
371 *Movement Ecology*, 8, 31.
- 372 Jonsen, I., Flemming, J., & Myers, R. (2005). Robust state-space modeling of animal move-
373 ment data. *Ecology*, 86(11), 2874–2880.
- 374 Kays, R., Crofoot, M. C., Jetz, W., & Wikelski, M. (2015). Terrestrial animal tracking as an
375 eye on life and planet. *Science*, 348(6240).
- 376 Kristensen, K., Nielsen, A., Berg, C. W., Skaug, H., & Bell, B. M. (2016). TMB: Automatic
377 differentiation and Laplace approximation. *Journal of Statistical Software*, 70, 1–21.
- 378 Lobo, J. M., Jiménez-Valverde, A., & Hortal, J. (2010). The uncertain nature of ab-
379 sences and their importance in species distribution modelling. *Ecography*, 33(1), 103–114.
380 <https://doi.org/https://doi.org/10.1111/j.1600-0587.2009.06039.x>
- 381 Lopez, R., Malardé, J.-P., Danès, P., & Gaspar, P. (2015). Improving Argos Doppler location
382 using multiple-model smoothing. *Animal Biotelemetry*, 3, 32.
- 383 McClintock, B. T., & Michelot, T. (2018). MomentuHMM: R package for generalized hidden
384 Markov models of animal movement. *Methods in Ecology and Evolution*, 9, 1518–1530.
- 385 McMahon, C. R., Roquet, F., Baudel, S., Belbeoch, M., Bestley, S., Blight, C., Boehme, L.,
386 Carse, F., Costa, D. P., Fedak, M. A., Guinet, C., Harcourt, R., Heslop, E., Hindell, M.
387 A., Hoennner, X., Holland, K., Holland, M., Jaine, F. R. A., Jeanniard du Dot, T., ...
388 Woodward, B. (2021). Animal borne ocean sensors – AniBOS – an essential component
389 of the global ocean observing system. *Frontiers in Marine Science*, 8, 751840.
- 390 Michelot, T., Langrock, R., Bestley, S., Jonsen, I. D., Photopoulou, T., & Patterson, T.
391 A. (2017). Estimation and simulation of foraging trips in land-based marine predators.
392 *Ecology*, 98(7), 1932–1944.
- 393 Michelot, T., Langrock, R., & Patterson, T. A. (2016). MoveHMM: An R package for the
394 statistical modelling of animal movement data using hidden Markov models. *Methods in
395 Ecology and Evolution*, 7, 1308–1315.
- 396 Patterson, T. A., Parton, A., Langrock, R., Blackwell, P. G., Thomas, L., & King, R. (2017).
397 Statistical modelling of individual animal movement: An overview of key methods and a
398 discussion of practical challenges. *AStA Advances in Statistical Analysis*, 101, 399–438.
- 399 Patterson, T. A., Thomas, L., Wilcox, C., Ovaskainen, O., & Matthiopoulos, J. (2008).
400 State-space models of individual animal movement. *Trends in Ecology and Evolution*, 23,

- 401 87–94.
- 402 Phillips, L., Carroll, G., Jonsen, I. D., Harcourt, R. G., Brierley, A., Wilkins, A., & Cox, M.
403 (2021). Variability in prey field structure drives inter-annual differences in prey encounter
404 by a marine predator, the little penguin. *Proceedings of the Royal Society B, In Review*.
- 405 Phillips, S. J., Dudík, M., Elith, J., Graham, C. H., Lehmann, A., Leathwick, J., & Ferrier,
406 S. (2009). Sample selection bias and presence-only distribution models: implications for
407 background and pseudo-absence data. *Ecological Applications*, 19(1), 181–197. <https://doi.org/10.1890/07-2153.1>
- 408 R Core Team. (2021). *R: A language and environment for statistical computing*. R Foundation
409 for Statistical Computing. <https://www.R-project.org/>
- 410 Raymond, B., Lea, M.-A., Patterson, T., Andrews-Goff, V., Sharples, R., Charrassin, J.-
411 B., Cottin, M., Emmerson, L., Gales, N., Gales, R., Goldsworthy, S. D., Harcourt, R.,
412 Kato, A., Kirkwood, R., Lawton, K., Ropert-Coudert, Y., Southwell, C., Hoff, J. van den,
413 Wienecke, B., ... Hindell, M. A. (2015). Important marine habitat off east Antarctica
414 revealed by two decades of multi-species predator tracking. *Ecography*, 38, 121–129.
- 415 Schick, R. S., New, L. F., Thomas, L., Costa, D. P., Hindell, M. A., McMahon, C. R.,
416 Robinson, P. W., Simmons, S. E., Thums, M., J., Harwood., & Clark, J. S. (2013).
417 Estimating resource acquisition and at-sea body condition of a marine predator. *Journal
418 of Animal Ecology*, 82, 1300–3015.
- 419 Service Argos. (2016). *Argos Users' Manual*. CLS. <http://www.argos-system.org/manual>
- 420 Sumner, M. D., Wotherspoon, S. J., & Hindell, M. A. (2009). Bayesian estimation of animal
421 movement from archival and satellite tags. *PLoS ONE*, 4(10). <https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0007324>
- 422 Thygesen, U. H., Albertsen, C. M., Berg, C. W., Kristensen, K., & Neilsen, A. (2017). Vali-
423 dation of ecological state space models using the Laplace approximation. *Environmental
424 and Ecological Statistics*. <https://doi.org/DOI: 10.1007/s10651-017-0372-4>
- 425 Whoriskey, K., Auger-Méthé, M., Albertsen, C. M., Whoriskey, F. G., Binder, T. R., Kreuger,
426 C. C., & Mills Flemming, J. (2017). A hidden markov movement model for rapidly
427 identifying behavioral states from animal tracks. *Ecology and Evolution*, 7, 2112–2121.