

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

<sup>§</sup> Ian D. Jonsen<sup>1,\*</sup>, W. James Grecian<sup>2,a</sup>, Lachlan Phillips<sup>1</sup>, Gemma Carroll<sup>3</sup>, Clive  
<sup>6</sup> McMahon<sup>4</sup>, Robert G. Harcourt<sup>1</sup>, Mark A. Hindell<sup>5</sup>, and Toby A. Patterson<sup>6</sup>

<sup>7</sup> <sup>1</sup>School of Natural Sciences, Macquarie University, Sydney, NSW, Australia

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

<sup>10</sup>Environmental Defense Fund, Seattle, WA, United States

<sup>11</sup> <sup>4</sup>Sydney Institute of Marine Science, Mosman, NSW, Australia

<sup>12</sup> <sup>5</sup>Institute for Marine and Antarctic Studies, University of Tasmania, Hobart, T

<sup>13</sup> <sup>6</sup>CSIRO Oceans and Atmosphere, Hobart, TAS, Australia

<sup>14</sup>\*corresponding author, ian.jonsen@mq.edu.au

<sup>15</sup> <sup>a</sup>Present address: Department of Geography, Durham University, Dur

16 Kingdom

17 **Abstract**

- 18 1. Animal tracking data are indispensable for understanding the ecology, behaviour and  
19 physiology of mobile or cryptic species. Meaningful signals in these data can be obscured  
20 by noise due to imperfect measurement technologies, requiring rigorous quality control  
21 as part of any comprehensive analysis.
- 22 2. State-space models are powerful tools that can separate meaningful signals from noise.  
23 These tools are ideal for quality control of error-prone location data and for inferring  
24 where animals are and what they are doing when they record or transmit other informa-  
25 tion. However, these statistical models can be challenging and time-consuming to fit to  
26 diverse animal tracking data sets.
- 27 3. The R package **foieGras** eases the tasks of conducting quality control on and inference  
28 of changes in movement from animal tracking data. This is achieved via: 1) a simple but  
29 extensible workflow that accommodates both novice and experienced users; 2) automated  
30 processes that alleviate complexity from data processing and model specification/fitting  
31 steps; 3) simple movement models coupled with a powerful numerical optimization  
32 approach for rapid and reliable model fitting.
- 33 4. We highlight **foieGras**' capabilities through three applications to real animal tracking  
34 data. Full R code for these and additional applications are included as Supporting  
35 Information so users can gain a deeper understanding of how to use **foieGras** for their  
36 own analyses.

37 **Keywords:**

- 38 animal movement, biologging, bio-telemetry, movement behaviour, move persistence, random  
39 walk, simulation, state-space model

## 40 1 | Introduction

41 The collection of animal telemetry and bio-logging data has become essential for understanding  
42 the movements, behaviour, social interactions, foraging ecology, physiology, habitat use and  
43 population dynamics of mobile and/or cryptic species (Kays *et al.*, 2015). In addition, the  
44 sophistication and precision of current technology now enables the use of tagged animals to  
45 collect in situ environmental information that complements data collected from traditional  
46 Earth observing platforms (McMahon *et al.*, 2021). However, in each of these applications,  
47 common issues with tracking data such as irregularly timed measurements, sensor biases and  
48 location measurement errors must be overcome.

49 State-space models (SSMs) are powerful tools for conducting quality control of and mak-  
50 ing behavioural inference from animal tracking data (Patterson *et al.*, 2008). Within this  
51 framework, an animal's true location and/or behaviour in the wild are unobserved states,  
52 while measurements recorded by animal-borne sensors provide noisy observations. The SSM  
53 comprises both a movement process model (e.g., a correlated random walk) and a measurement  
54 model that are fit to animal tracks to estimate true location and behaviour states, thereby  
55 separating ecologically-meaningful signals from the noise of sensor measurements.

56 Here we introduce **foieGras** (pronounced “*fwah grah*”), a package for fitting SSMs in R (R  
57 Core Team, 2021). This package has two aims: (1) to be a simple and fast implementation  
58 of SSMs to quality control error-prone animal location data (Jonsen *et al.*, 2020); and (2)  
59 for inference of changes in movement behaviour along animal tracks (Jonsen *et al.*, 2019).  
60 Although several implementations of SSMs for animal movement data exist, e.g. **bsam** (Jonsen  
61 *et al.*, 2005), **crawl** (Johnson *et al.*, 2008), and **ctmm** (Calabrese *et al.*, 2016), their broad  
62 utility can be hampered by their user interfaces that tend to be technically complex. **foieGras**  
63 has a more user-friendly implementation suitable for novices, yet users can exert control over  
64 many aspects of the package functions via optional arguments, accommodating the needs of  
65 experienced users.

66 We describe the main features of **foieGras** and illustrate its use through a set of applications  
67 drawing on Argos and GPS tracking data. Full R code and data for each of the applications  
68 is provided in the Supporting Information. Additional details on package functions and use  
69 can be found in the package help files and vignettes.

70 **2 | foieGras overview**

71 The workflow for **foieGras** is deliberately simple, with many of the usual track data processing  
72 checks handled automatically. Here we outline the main aspects of the **foieGras** package.  
73 The package's main functions with brief code examples are listed in Table 1, and a generalized  
74 workflow with example code is depicted in Figure 1.

75 **2.1 | Data input - format\_data**

76 Animal tracking data, consisting of a time-series of location coordinates, can be read into R as  
77 a **data.frame** or **tibble** using standard functions such as **read.csv** or **read\_csv**. Possible  
78 default input data formats are shown in Figure 1. The **format\_data()** function can be used  
79 to coerce non-default data formats into that expected by **fit\_ssm()**. **format\_data()** can be  
80 used explicitly in a workflow or it will be called automatically by **fit\_ssm()**. More details  
81 on input data formats and the use of **format\_data()** are provided in the Overview vignette  
82 (<https://ianjonsen.github.io/foieGras/articles/Overview.html>). Additionally, **foieGras** can  
83 accept data as an **sf-tibble** or **sf data.frame** with locations in any recognized projection,  
84 in which case the SSM-estimated locations will be returned in the same projection.

Table 1: Main `foieGras` functions with description and example code. In the code examples, `...` denotes optional arguments. Details on these arguments can be found in the package help files for each function (e.g., `?fit_ssm()`).

Function	Description	Code
<code>format_data()</code>	Coerce tracking data into the format expected by <code>'fit_ssm()'</code>	<code>data &lt;- format_data(x, ...)</code>
<code>fit_ssm()</code>	Fit a State-Space Model to location data	<code>fit &lt;- fit_ssm(data, model, time.step, ...)</code>
<code>fit_mpm()</code>	Fit a Move Persistence Model to location data	<code>fmp &lt;- fit_mpm(fit, model, ...)</code>
<code>grab()</code>	Extract fitted/predicted/observed locations from a foieGras model, with or without projection information	<code>locs &lt;- grab(fit, what = 'predicted', ...)</code>
<code>osar()</code>	Estimate One-Step-Ahead Residuals from a foieGras SSM	<code>res &lt;- osar(fit, ...)</code>
<code>map()</code>	Map fitted/predicted locations with or without a defined projection	<code>map(fit, fmp, what = 'predicted', ...)</code>
<code>sim()</code>	Simulate individual animal tracks with Argos LS or KF errors	<code>slocs &lt;- sim(N, model, ...)</code>
<code>sim_fit()</code>	Simulate animal tracks from SSM model fit objects	<code>sfit &lt;- simfit(fit, ...)</code>
<code>sim_filter()</code>	Filter simulated tracks according to similarity criteria	<code>sfit &lt;- sim_filter(sfit, keep, ...)</code>
<code>route_path()</code>	Reroute path so estimated locations are off land	<code>fit &lt;- route_path(fit, what = 'predicted', ...)</code>
<code>plot.ssm_df()</code>	Plot the fit of a foieGras SSM to data	<code>plot(fit, what = 'predicted', ...)</code>
<code>plot.osar()</code>	Plot One-Step-Ahead Residuals from a foieGras SSM	<code>plot(res, type = 'ts')</code>

85 **2.2 | State-space model fitting - `fit_ssm()`**

86 State-space models are fit using `fit_ssm()` (Fig. 1). Data type is detected from the location  
87 quality classes typical of Argos data; classes can be added for other data types. Based on  
88 the location quality classes and optional information on measurement errors contained in the  
89 data, an appropriate measurement error model is selected for each observation (Jonsen *et al.*,  
90 2020). Fits to single versus multiple individuals are handled automatically, with sequential  
91 fitting for the latter. No hierarchical or pooled estimation among individuals is currently  
92 possible. `foieGras`' process and measurement error models are outlined in Table 2.

93 Many optional arguments can be specified in `fit_ssm()`, and are explained in the package  
94 documentation. We focus only the essential ones here:

- 95     • `data` the input data structured as illustrated in Fig. 1  
96     • `model` the process model to be used (one of `rw`, `crw`, or `mp`; see Table 2 for details)  
97     • `time.step` the prediction time interval (expressed in hours)

98 The function invokes an automated data processing stage where the data type and measurement  
99 error model(s) are determined, observation times are sorted and checked for duplicates, and a  
100 speed filter identifies extreme locations to be ignored by the SSM. The SSM is then fitted  
101 to the processed data, with the user-specified process model and automatically selected  
102 measurement error model(s). The R package `TMB`, Template Model Builder (Kristensen  
103 *et al.*, 2016), is used to compute the gradient function in C++, which greatly speeds model  
104 fitting. A fit object is returned as a nested `tibble`, listing the individual animal id(s), basic  
105 convergence information and a list of model output including estimated parameters and states,  
106 processed data, and diagnostic information. A `data.frame` of SSM fitted (location estimates  
107 corresponding to the observation times) or predicted values (locations predicted at regular  
108 `time.step` intervals) can be extracted using the `grab()` function. Parameter estimates, AIC  
109 and other information are viewed using the `summary()` function. Further details on SSM  
110 fitting are here: ([https://ianjonsen.github.io/foieGras/articles/SSM\\_fitting.html](https://ianjonsen.github.io/foieGras/articles/SSM_fitting.html)).

Table 2: `foieGras` movement process and measurement error models with brief descriptions and parameters. The process and measurement error models are the two components of a state-space model. Parameters listed are those estimated when fitting the state-space model to data, these estimates can be accessed via `summary()` (see Overview vignette for details).

Model	Description	Parameters
<b>Process Models</b>		
<code>rw</code> - Random walk	Movements are random in direction and magnitude.	$\rho_p, \sigma_x, \sigma_y$
<code>crw</code> - Correlated random walk	Movements are random and correlated in direction and magnitude.	D
<code>mp</code> - Move persistence	Movements are random with correlation in direction and magnitude that varies in time.	$\rho_p, \sigma_x, \sigma_y, \sigma_g$
<b>Measurement Error Models</b>		
<code>LS</code> - Argos least-squares	Location measurement errors assigned via Argos location quality class and error multiplication factors (see <code>?emf</code> for details).	$\rho_o, \tau_x, \tau_y$
<code>KF</code> - Argos Kalman filter	Location measurement errors specified by error ellipse variables ( <code>smaj</code> , <code>smin</code> , <code>eor</code> ) present in data (see Overview vignette for details).	$\psi$
<code>G</code> - GPS	Location measurement errors are assumed by default to be 10% of Argos location quality class 3 (see <code>?emf</code> for details).	$\rho_o, \tau_x, \tau_y$
<code>GL</code> - Geolocation	Location measurement errors are specified by <code>lonerr</code> , <code>laterr</code> variables present in data.	$\rho_o, \tau_x, \tau_y$

$\rho_p$  - correlation between x and y process variances (`rho_p`)

$\sigma_x, \sigma_y$  - process standard deviation in x and y directions, respectively (`sigma_x, sigma_y`)

D - process variance in x and y directions (D)

$\sigma_g$  - standard deviation of random walk on  $\gamma_t$  (`sigma_g`)

$\rho_o$  - correlation between x and y measurement error variances (`rho_o`)

$\tau_x, \tau_y$  - standard deviation of measurement errors in x and y directions, respectively (`tau_x, tau_y`)

$\psi$  - scaling parameter to account for potential bias in semi-minor axis of Argos KF error ellipses (`psi`)

111 **2.3 | Behavioural estimation - `fit_ssm()`, `fit_mpm()`**

112 Move persistence, an index of movement behaviour, can be estimated as a continuous-valued  
113 ( $0 - 1$ ), time-varying latent variable that represents changes in movement pattern based  
114 on autocorrelation in speed and direction (Jonsen *et al.*, 2019). There are two approaches  
115 in `foieGras` for estimating move persistence. The first is to use `fit_ssm()` with `model =`  
116 '`mp`', which fits a continuous-time move persistence model in state-space form and thereby  
117 simultaneously estimates true locations and move persistence from the error-prone telemetry  
118 data (Fig. 1). This approach is most appropriate for fitting to irregularly-timed and error-  
119 prone Argos data as both aspects are taken into account explicitly. The second is to use  
120 `fit_mpm()`, which can take as input either location data or SSM-estimated locations from  
121 an `fit_ssm()` model fit object. This approach is generally more appropriate when the data  
122 have minimal measurement error (e.g., GPS locations; see Application 3.2). Further details  
123 on fitting move persistence models are here: [https://ianjonsen.github.io/foieGras/articles/Move\\_persistence\\_models.html](https://ianjonsen.github.io/foieGras/articles/Move_persistence_models.html).

125 **2.4 | Model checking and visualization - `osar()`, `plot()`, `map()`**

126 Before using fitted or predicted locations, a `fit_ssm()` model fit should be checked and  
127 visualized to confirm that the model adequately describes the data. The `osar()` function  
128 computes one-step-ahead (prediction) residuals via the `oneStepPredict()` function from the  
129 TMB R package (Kristensen *et al.*, 2016; Thygesen *et al.*, 2017). A set of residuals are calculated  
130 for the `x` and `y` values corresponding to the fitted values from the SSM. A generic `plot()`  
131 (`plot.osar`) method provides an easy way to visualize the prediction residuals as time-series  
132 plots, quantile-quantile plots, or autocorrelation functions (Fig. 1). Further details on model  
133 validation are here: [https://ianjonsen.github.io/foieGras/articles/SSM\\_validation.html](https://ianjonsen.github.io/foieGras/articles/SSM_validation.html).

134 State-space model fits to data can also be visualised by using the generic `plot()` (`plot.ssm_df`)  
135 function on a model fit object. Options exist to plot fitted or predicted values along with  
136 observations as either paired, 1-D time-series (using the `type = 1` argument), or as 2-D tracks

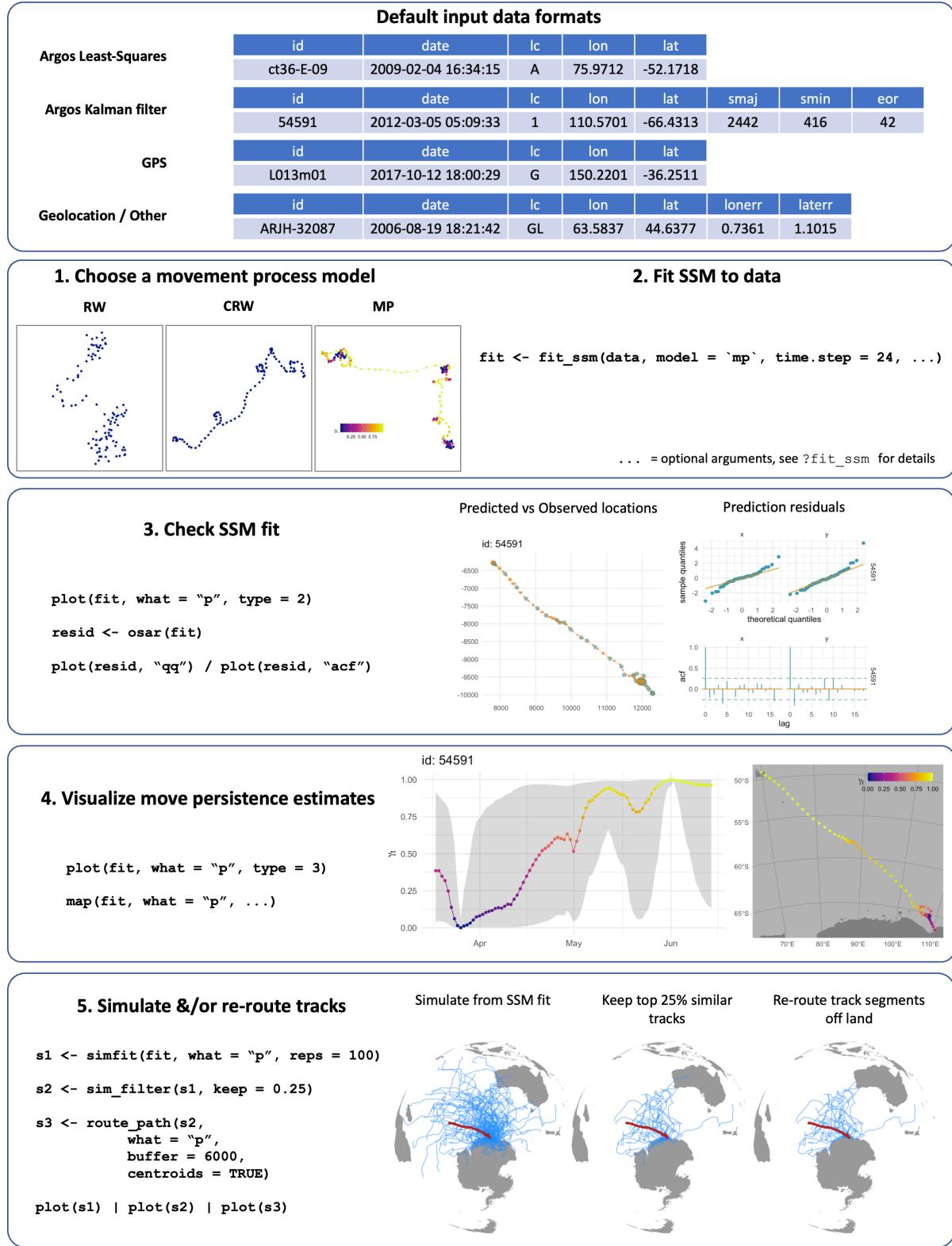


Figure 1: A generalized `foieGras` workflow showing default input data formats, model fitting, model checking/validation, visualisation of movement behaviour estimates, track simulation and re-routing around land. Example code is displayed for each stage. Details on default input data formats and options for handling other formats are provided in the Overview vignette.

`137 (type = 2)` with 95% confidence intervals or ellipses. These plots provide a rapid check on  
138 SSM fits to data. Additionally, when the fitted SSM is the move persistence model (i.e.,  
139 `model = 'mp'`), 1-D time-series (`type = 3`) or 2-D track plots (`type = 4`) of move persistence  
140 estimates can be displayed (Fig. 1).

`141 fit_ssm()` model fits can be mapped using `map()` for single or multiple individuals. By  
142 default, `map()` uses the coastline data from the `rnatuelearth` R package (South, 2022)  
143 at medium or high resolution (Fig. 1), but can also use tiled maps for finer-scale detail,  
144 via the `rosm` (Dunnington, 2019) and `ggspatial` (Dunnington, 2021) R packages. Mapping  
145 aesthetics (e.g., plot symbols, sizes, colours, fills) can be customized via the `aes` argument  
146 and use of the `aes_1st()` function. See code in SI for examples. Further details on mapping  
147 tracks are here: <https://ianjonsen.github.io/foieGras/articles/Mapping.html>.

`148 All foieGras visualizations draw on the ggplot2 R package (Wickham, 2016), with multi-`  
149 `panel plots also using the patchwork R package (Pedersen, 2020), and generally can be`  
150 `modified through additive calls in the usual ggplot2 manner. See code in SI for examples.`

## `151 2.5 | Simulation - sim(), sim_fit(), sim_filter()`

`152 Track simulation can be a helpful, yet informal, way of evaluating the degree to which`  
153 `statistical movement models capture essential features of animal movement data (Michelot`  
154 `et al., 2017). A variety of movement processes are simulated using sim(), including the rw,`  
155 `crw, and mp process models, as well as rw and crw models with switching between multiple`  
156 `movement states that are not available in foieGras's SSM's. sim() is intended as a general`  
157 `tool for random track simulation for a variety of purposes such as qualitative comparisons to`  
158 `real data, or evaluation of SSM estimation biases.`

`159 The sim_fit() function is intended for random track simulation in support of habitat usage`  
160 `and similar analyses. sim_fit() extracts movement parameters from an SSM fit object and`  
161 `simulates random tracks based on these parameters and with the same duration as the SSM-`

162 estimated track. `sim_fit()` has optional arguments that impose additional constraints on the  
163 simulated tracks that would be too computationally challenging to incorporate into the SSM.  
164 Whereas the SSM is constrained by the data, the simulated tracks are relatively unconstrained  
165 and so the following arguments allow the simulations to more closely approximate the real  
166 tracks.

167 The argument `cpf = TRUE` allows simulation of central place foragers by ensuring that tracks  
168 start and end at approximately the same location. Movements can also be constrained to  
169 remain mostly in water via a potential function (Preisler *et al.*, 2013), using gradient rasters  
170 included in `foieGras` and the `grad` and `beta` arguments. These are illustrated in the code  
171 for Application 3.3.

172 The `sim_filter()` function provides a tool to filter the simulated tracks based on their  
173 similarity to the SSM-estimated track. Here, similarity is calculated as the sum of normalized  
174 difference in net displacement (km) and overall bearing (deg) between the SSM-estimated  
175 track and the simulated tracks. Filtering uses one of two metrics that capture the differences  
176 between the two tracks (see `?similarity_flag` for more detail). These metrics are motivated  
177 by the ‘flag value’ described in Hazen *et al.* (2017). The the quantile of flag values to be retain  
178 is specified via the `keep` argument; i.e. `keep = 0.25` (the default) will return a `sim_fit`  
179 object containing those simulated tracks with flag values in the top 25% of values calculated  
180 for the input `sim_fit` object (Fig. 1).

181 Further details on track simulation are here: <https://ianjonsen.github.io/foieGras/articles/Tr>  
182 ack\_simulation.html.

183 **2.6 | Path rerouting - `route_path()`**

184 As the SSMs implemented in `foieGras` have no information about potential barriers to animal  
185 movement it is possible for locations to be estimated in implausible locations, such as on  
186 land for marine species. To overcome this, `foieGras` makes use of the `pathroutr` R package

187 (London, 2020) to efficiently re-route locations from land back to water by using visibility  
188 graphs (Jan *et al.*, 2014). `route_path()` can be applied to either an SSM fit object or the  
189 simulations generated by `sim_fit()`. Users must apply `route_path()` themselves, this is not  
190 done automatically in `foieGras`. When the input is an SSM fit object the re-routed path can  
191 be appended to the object for visualization and use in subsequent analyses. When the input  
192 is a `sim_fit` object the locations within the simulation are replaced with the re-routed paths  
193 (Fig. 1). We illustrate the latter in Application 3.3. Further details on path rerouting are  
194 here: [https://ianjonsen.github.io/foieGras/articles/Path\\_rerouting.html](https://ianjonsen.github.io/foieGras/articles/Path_rerouting.html).

## 195 3 | Applications

196 We illustrate the main capabilities of `foieGras` with three applications. Complete R code for  
197 reproducing the applications and for gaining a deeper understanding of `foieGras` functions are  
198 provided in the Supporting Information (Appendices S1 - S4). Data used in the applications  
199 are available on Dryad.

### 200 3.1 | SSM validation with prediction residuals

201 We use a sub-adult male southern elephant seal (*Mirounga leonina*) track included as example  
202 data in `foieGras` (`sese2`, id: ct36-E-09), sourced from the Australian Integrated Marine  
203 Observing System (IMOS; data publicly available via imos.aodn.org.au) deployments at Iles  
204 Kerguelen in collaboration with the French IPEV and SNO-MEMO programmes. The data  
205 are temporally irregular Argos Least-Squares locations. We fitted both the `rw` and `crw` models  
206 using `fit_ssm()` with a speed filter threshold (`vmax`) of 4 ms<sup>-1</sup> and a 12-h time step. We  
207 calculated prediction residuals using `osar()`, and then plotted the `osar` residuals to assess  
208 and compare the model fits (Fig. 2).

209 Plots of predicted states with the observations suggest both models yield similar global fits  
210 (Fig. 2a; compare blue versus red lines and 95% confidence ellipses), but discrepancies are

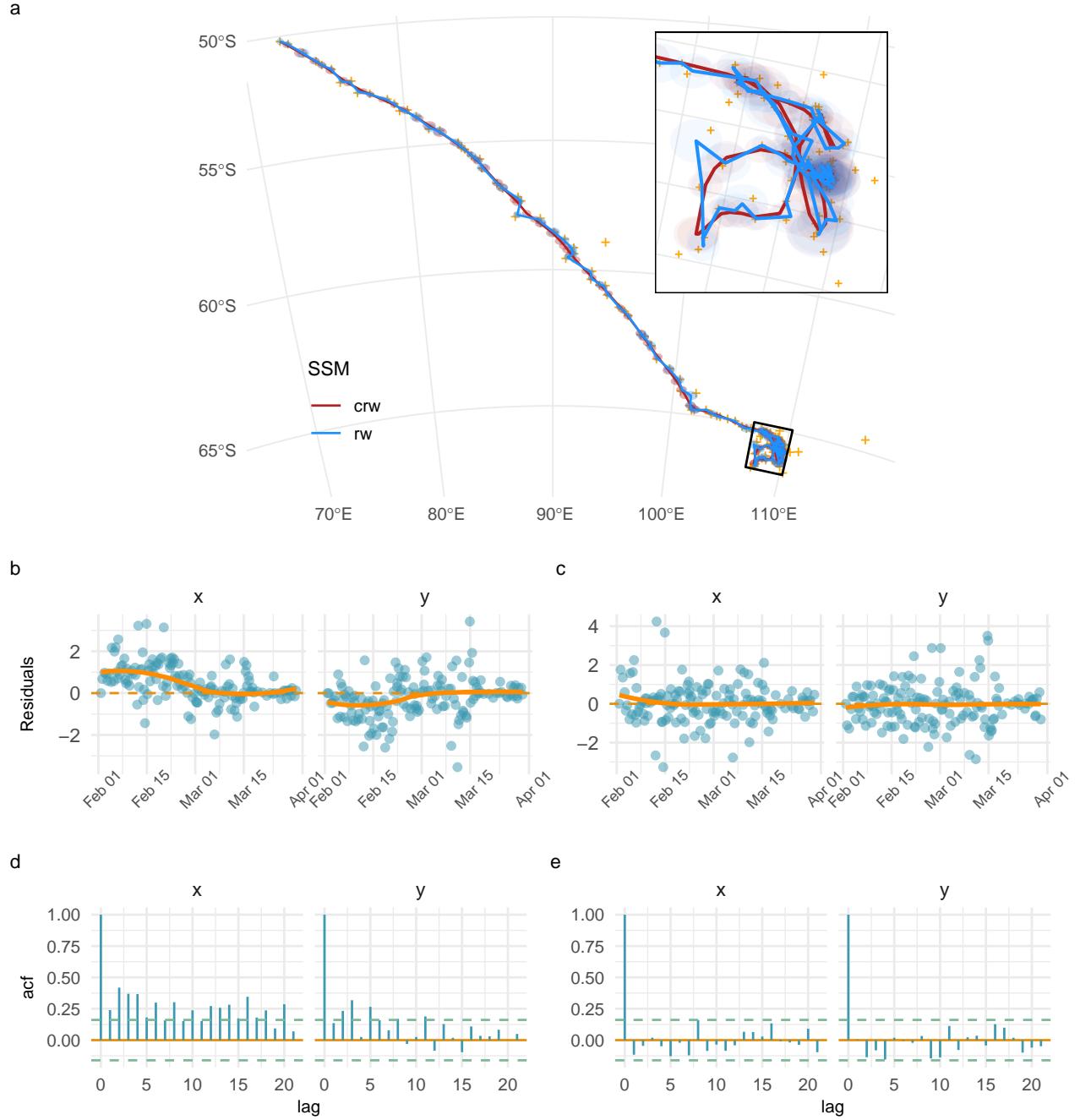


Figure 2: State-space model fits to a southern elephant seal track with finer scale detail inset (a), and diagnostic plots for assessing goodness-of-fit of the **rw** (b - prediction residual time-series; d - prediction residual autocorrelation) and **crw** (c,e) state-space models. All residual plots generated using the `plot.osar` function.

visible at a finer scale (Fig. 2a, inset). Predicted locations from the two models differ by a median 6.62 km (range: 0.02, 53.02 km), and there are marked trends in the residuals for the `rw` model fit (Fig. 2b) with significantly positive autocorrelation in both the x and y directions (Fig. 2d). The `crw` prediction residuals show no temporal trend and have little autocorrelation (Fig. 2c,e), implying the `crw` process model is a better fit to the data.

Although the `crw` process model provides a better fit to the data, confidence ellipses around fitted and predicted location estimates should be examined. The `plot.ssm_df` function provides an easy way to visualise estimated location uncertainty as 95% confidence intervals/ellipses in 1 (`type = 1`) or 2 dimensions (`type = 2`). When confidence ellipses are large, use of the location estimates without acknowledging their uncertainty in subsequent analyses may lead to biased results.

### 3.2 | Inferring move persistence

To illustrate how move persistence ( $\gamma_t$ ) can be estimated from GPS data, we use four little penguin (*Eudyptula minor*) tracks of daily foraging trips during the chick-rearing period from Montague Island, NSW, Australia (described in Phillips *et al.*, 2022). The data are temporally irregular GPS locations, with high frequency sampling (mean 15 s) intermittently disrupted by the birds' diving behaviour, and are assumed to have minimal measurement error. We fitted the `crw` SSM to the GPS data to predict temporally regular locations at 2-min intervals, and assumed consistently small bivariate normal location measurement errors (ie.  $\pm 30$  m sd). We then used `fit_mpm` to estimate  $\gamma_t$  from these regularized locations. As  $\gamma_t$  estimates are sensitive to time interval choice, we compare 2-min estimates with those from both shorter and longer prediction intervals. Finally, we visually compared  $\gamma_t$  estimates with prey capture events inferred from accelerometry data (see Carroll *et al.*, 2014, for details) to ascertain whether reduced  $\gamma_t$  is indicative of active foraging by these penguins. We provide an extra application to illustrate how move persistence can be inferred from Argos data using `fit_ssm` (Appendix S3).

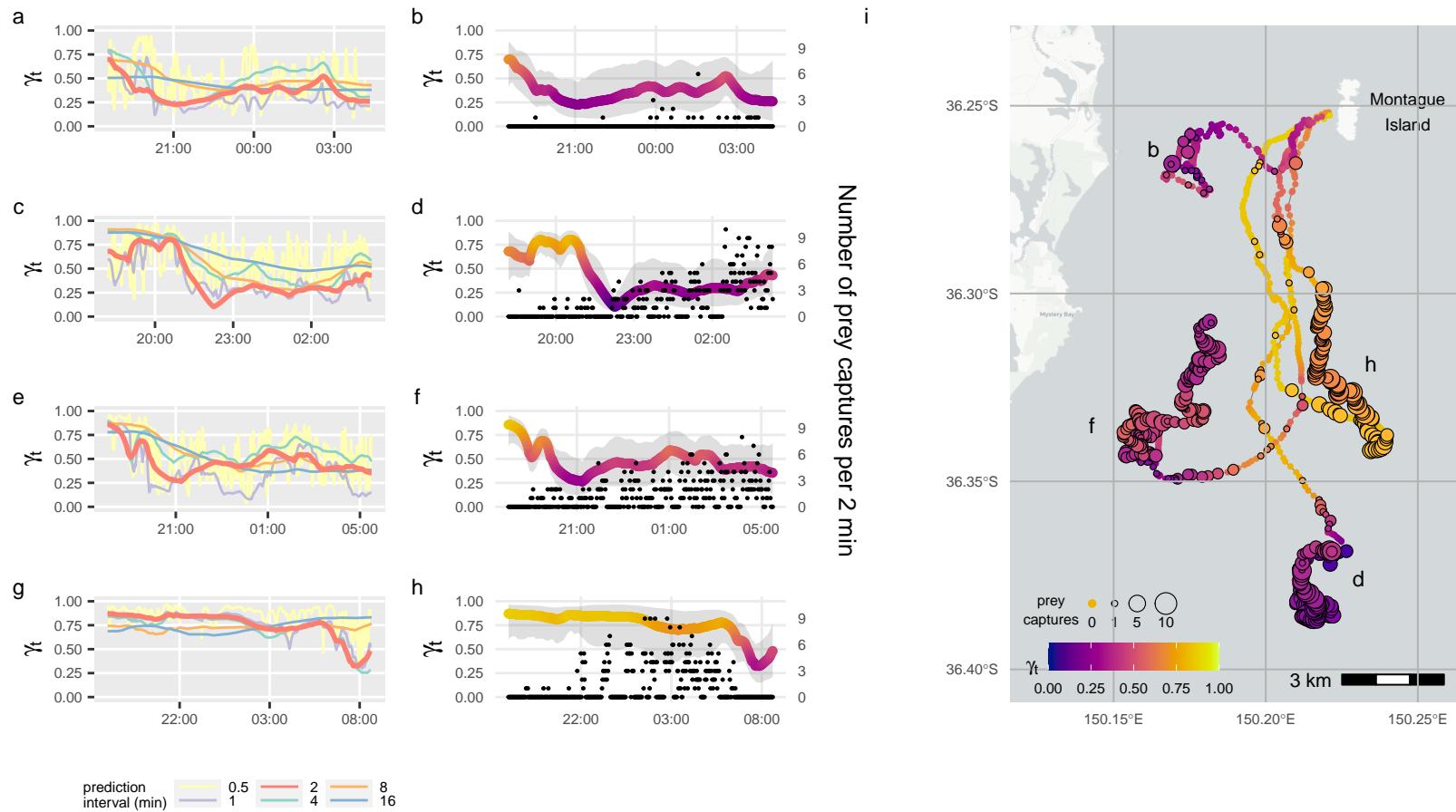


Figure 3: Inferred move persistence,  $\gamma_t$ , from time-regularised little penguin GPS tracks.  $\gamma_t$  estimates were compared across a range of prediction intervals (0.5 to 16 min; a,c,e,g). The 2-min  $\gamma_t$  estimates for each penguin are overlaid with the number of prey capture events, inferred from accelerometry data (black circles in b,d,f,h; grey envelopes are 95 % CI's) and along SSM-predicted penguin tracks (i). Location symbol sizes in (i) are proportional to the number of prey capture events at 2-min intervals corresponding to the SSM predicted location times.

237 As  $\gamma_t$  is sensitive to choice of prediction time interval, it is important that a range of intervals  
238 be evaluated. Although the chosen 2-min prediction interval is longer than the GPS sampling  
239 frequency (15 s), it adequately captures the main changes in  $\gamma_t$  (Fig. 3a,c,e,g). Shorter  
240 intervals (0.5 and 1 min) have substantial short-term variability that may not reflect real  
241 changes in movement pattern, while longer intervals (8 and 16 min) likely smooth out too  
242 much real variability in the movements.

243 The little penguin GPS tracks did not exhibit strong contrast in move persistence ( $\gamma_t$  range  
244 = 0.102, 0.871; Fig. 3b,d,f,h). Nonetheless, the move persistence estimates highlight changes  
245 in movement pattern over the course of the penguins' daily foraging trips (Fig. 3i). The  
246 penguins departed Montague Is. with relatively fast movements, three directed southward  
247 (Fig. 3 tracks d,f,h in i) and one less directed and remaining closer to the island (track b),  
248 before slowing down and engaging in meandering movements (orange - purple in Fig. 3i).  
249 The high  $\gamma_t$  estimates at the start of the penguins' foraging trips were associated with few  
250 prey captures (Fig. 3b,d,f). However, periods of lower  $\gamma_t$  were only loosely associated with  
251 higher prey captures. For example, penguin "h" had no prey captures as  $\gamma_t$  declined toward  
252 the end of its foraging trip (Fig. 3h). This period of lower  $\gamma_t$  is likely indicative of resting  
253 as returning penguins usually wait offshore until after dusk before hauling out on the island.  
254 These results highlight that while move persistence can objectively indicate where animals  
255 spend more (low  $\gamma_t$ ) or less (high  $\gamma_t$ ) time, the behaviour(s) responsible for these patterns  
256 may not be obvious or consistent (e.g., Riaz *et al.*, 2021).

### 257 3.3 | Simulating tracks from `foieGras` model fits

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

263 We simulate 50 animal movement paths from the `crw` process model using `sim_fit`, and apply  
 264 a potential function using the `grad` and `beta` arguments to constrain the simulated paths  
 265 to largely remain in water. These tracks are then filtered based on their similarity to the  
 266 original path using `sim_filter` and the top 25 % retained (`keep = 0.25`)(Fig. 4a,b). As the  
 267 potential function does not guarantee all locations remain off land, we re-route any remaining  
 268 simulated locations from land back to water using `route_path` (Fig. 4c). In combination,  
 269 these functions provide a pragmatic, non-statistical method to generate and objectively filter  
 270 pseudo-tracks for use in movement or habitat modelling applications.

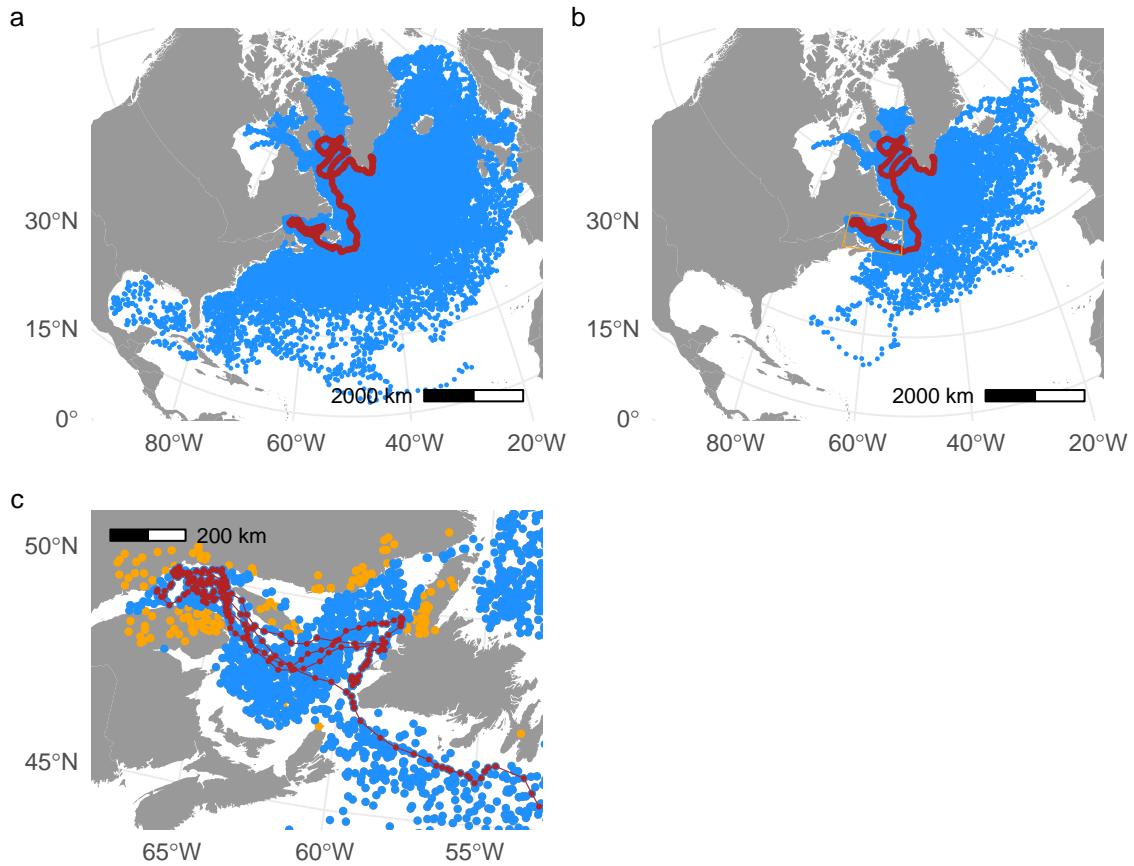


Figure 4: Simulating (a) 100 movement paths from a correlated random walk process model; (b) filtering those tracks to select the top 25% based on their similarity to the original SSM-predicted track (red); and (c) re-routing simulated locations on land (orange) back to ocean (blue). The orange box in (b) indicates region magnified in (c). SSM-predicted track (red) overlaid in all panels for context.

271 **4 | Conclusions**

272 The **foieGras** package was developed to ease fitting state-space models for quality control  
273 of animal location data and for inference of behavioural change along animal tracks. We  
274 achieve these primarily through a simple yet extensible workflow, simple movement models,  
275 and computational speed. Combined, these traits accommodate both novice or occasional and  
276 advanced users, and facilitate use in automated, operational quality-assurance/quality-control  
277 processes for animal-borne sensor data (e.g., the AniBOS network McMahon *et al.*, 2021).

278 The **foieGras** package is an intermediate analysis toolbox where location quality control  
279 typically occurs after some initial data processing but prior to any comprehensive, final analysis.  
280 Similarly, the move persistence models rapidly identify changes in movement behaviour along  
281 animal tracks without any required *a priori* knowledge about the kind or number of behavioural  
282 states potentially hidden within the data. Subsequent, detailed behavioural analysis could  
283 entail use of a hidden Markov model to infer potential drivers of observed movements (e.g.,  
284 using the ‘moveHMM’ or ‘momentuHMM’ R packages Michelot *et al.*, 2016; McClintock &  
285 Michelot, 2018). Alternatively, move persistence - covariate relationships among multiple  
286 individuals can be inferred rapidly in a mixed-effects model framework using the **mpmm** R  
287 package (Jonsen *et al.*, 2019, <https://github.com/ianjonsen/mpmm>).

288 The **foieGras** track simulation tools support habitat usage analyses, allowing habitat avail-  
289 ability (Aarts *et al.*, 2012) to be estimated by generating ‘background’ points representing a  
290 null model of the distribution of foraging animals in the absence of external drivers (Raymond  
291 *et al.*, 2015). As the generation of such null locations can impact the performance of habitat  
292 usage models (Phillips *et al.*, 2009), **foieGras** allows simulated tracks to be filtered based on  
293 their similarity to the SSM-predicted tracks. Eventually, **foieGras** will also allow simulation  
294 from the **posterior** of SSM-predicted tracks, whereby the simulated tracks are constrained  
295 by the estimated locations and their uncertainty \citet[e.g., the **crawl** R package][]{Johnson  
296 *et al.*, 2008}.

297 **4.1 | The future of foieGras**

298 An older version of `foieGras` (v 0.7-6) is available on CRAN <https://CRAN.R-project.org/package=foieGras>, this version does not have all the features described here and  
299 will be archived by CRAN eventually. The current `foieGras` version (v 1.1) is available  
300 from the R-universe repository <https://ianjonsen.r-universe.dev/ui#package:foieGras>,  
301 and can be installed from source within R via `install.packages("foieGras", repos =`  
302 `"https://ianjonsen.r-universe.dev"`). Other installation options exist and we refer users  
303 to additional information here <https://github.com/ianjonsen/foieGras>.

304  
305 The `foieGras` package will undergo further development, expanding the range of available  
306 SSM's via new movement process models and/or enhanced measurement models for other  
307 electronic tracking data types. Additionally, we will seek to enhance integration with other  
308 R packages for processing and analysis of animal movement data, where feasible. Feedback  
309 from users is invaluable and encouraged. Users may submit bug reports and enhancement  
310 suggestions via the `foieGras` GitHub issues page (<https://github.com/ianjonsen/foieGras/issues/new/choose>). More general feedback is always welcome by contacting the lead author  
311 directly.  
312

313 **Acknowledgements**

314 IDJ acknowledges support from a Macquarie University co-Funded Fellowship and from  
315 partners: the US Office of Naval Research, Marine Mammal Program (grant N00014-18-1-  
316 2405); the Integrated Marine Observing System (IMOS); Taronga Conservation Society; the  
317 Ocean Tracking Network; Birds Canada; and Innovasea/VEMCO. TAP was supported by  
318 CSIRO Oceans & Atmosphere internal research funding scheme. The Integrated Marine  
319 Observing System (IMOS) supported seal fieldwork. IMOS is a national collaborative research  
320 infrastructure, supported by the Australian Government and operated by a consortium of  
321 institutions as an unincorporated joint venture, with the University of Tasmania as Lead

<sup>322</sup> Agent. Field work at Illes Kerguelen was conducted as part of the IPEV programme N°  
<sup>323</sup> 109 (PI H. WEIMERSKIRCH) and of the SNO-MEMO programme (PI C. GUINET) in  
<sup>324</sup> collaboration with IMOS. CTD tags were partly funded by CNES-TOSCA and IMOS. Little  
<sup>325</sup> penguin fieldwork was supported by an Australian Research Council Linkage grant to IDJ, GC  
<sup>326</sup> and RGH (LP160100162). Harp seal fieldwork was funded by Fisheries and Oceans Canada, a  
<sup>327</sup> bursary from Department for Business, Energy and Industrial Strategy (BEIS) administered  
<sup>328</sup> by the NERC Arctic Office, and the ARISE project (NE/P006035/1 and NE/P00623X/1), part  
<sup>329</sup> of the Changing Arctic Ocean programme jointly funded by the UKRI Natural Environment  
<sup>330</sup> Research Council (NERC) and the German Federal Ministry of Education and Research  
<sup>331</sup> (BMBF). All animal tagging procedures approved and executed under the Animal Ethics  
<sup>332</sup> Committee guidelines of the University of Tasmania (elephant seals), Macquarie University  
<sup>333</sup> (little penguins), and University of St Andrews (harp seals).

## <sup>334</sup> **Author's Contributions**

<sup>335</sup> IDJ developed the R package; WJG contributed harp seal data and to the R package; LP,  
<sup>336</sup> GC, and RGH contributed little penguin data; CRM, RGH and MAH contributed Southern  
<sup>337</sup> elephant seal data; IDJ and TAP developed the state-space models; IDJ wrote an initial  
<sup>338</sup> draft of the manuscript with contributions from WJG; all authors contributed to the final  
<sup>339</sup> manuscript.

## <sup>340</sup> **Data Accessibility**

<sup>341</sup> All code used here is provided in the `foieGras` package for R or in the Supporting Information.  
<sup>342</sup> Data will be made publicly available on Dryad here. For review purposes, the example  
<sup>343</sup> application data and associated .Rmd files can be accessed here: <https://github.com/ianjones>  
<sup>344</sup> en/foieGras.paper/tree/main/SI.

<sup>345</sup> **ORCID**

- <sup>346</sup> Ian D Jonsen <https://orcid.org/0000-0001-5423-6076>
- <sup>347</sup> W James Grecian <https://orcid.org/0000-0002-6428-719X>
- <sup>348</sup> Lachlan Phillips <https://orcid.org/0000-0002-7635-2817>
- <sup>349</sup> Gemma Carroll <https://orcid.org/0000-0001-7776-0946>
- <sup>350</sup> Clive R McMahon <https://orcid.org/0000-0001-5241-8917>
- <sup>351</sup> Robert G Harcourt <https://orcid.org/0000-0003-4666-2934>
- <sup>352</sup> Mark A Hindell <https://orcid.org/0000-0002-7823-7185>
- <sup>353</sup> Toby A Patterson <https://orcid.org/0000-0002-7150-9205>

<sup>354</sup> **References**

- <sup>355</sup> Aarts, G., Fieberg, J. & Matthiopoulos, J. (2012) Comparative interpretation of count,  
<sup>356</sup> presence-absence and point methods for species distribution models. *Methods in Ecology  
and Evolution*, **3**, 177–187.
- <sup>358</sup> Calabrese, J.M., Fleming, C.H. & Gurarie, E. (2016) ctmm: an R package for analyzing  
<sup>359</sup> animal relocation data as a continuous-time stochastic process. *Methods in Ecology and  
Evolution*, **7**, 1124–1132.
- <sup>361</sup> Carroll, G., Slip, D., Jonsen, I. & Harcourt, R. (2014) Supervised accelerometry analysis can  
<sup>362</sup> identify prey capture by penguins at sea. *Journal of Experimental Biology*, **217**, 4295–4302.
- <sup>363</sup> Dunnington, D. (2019) *rosm: Plot Raster Map Tiles from Open Street Map and Other Sources*.  
<sup>364</sup> R package version 0.2.5.
- <sup>365</sup> Dunnington, D. (2021) *ggspatial: Spatial Data Framework for ggplot2*. R package version  
<sup>366</sup> 1.1.5.
- <sup>367</sup> Grecian, W.J., Stenson, G.B., Biuw, M., Boehme, L., Folkow, L.P., Goulet, P.J., Jonsen, I.D.,  
<sup>368</sup> Malde, A., Nordøy, E.S., Rosing-Asvid, A. & Smout, S. (2022) Environmental drivers of

- 369 population-level variation in the migratory and diving ontogeny of an arctic top predator.
- 370 *Royal Society Open Science*, **9**.
- 371 Hazen, E.L., Palacios, D.M., Forney, K.A., Howell, E.A., Becker, E., Hoover, A.L., Irvine,  
372 L., DeAngelis, M., Bograd, S.J., Mate, B.R. & Bailey, H. (2017) WhaleWatch: a dynamic  
373 management tool for predicting blue whale density in the California Current. *Journal of*  
374 *Applied Ecology*, **54**, 1415–1428.
- 375 Jan, G.E., Sun, C.C., Tsai, W.C. & Lin, T.H. (2014) An  $o(n \log n)$  shortest path algorithm  
376 based on delaunay triangulation. *IEEE/ASME Transactions on Mechatronics*, **19**, 660–666.
- 377 Johnson, D.S., London, J.M., Lea, M. & Durban, J.W. (2008) Continuous-time correlated  
378 random walk model for animal telemetry data. *Ecology*, **89**, 1208–1215.
- 379 Jonsen, I.D., Patterson, T.A., Costa, D.P., Doherty, P.D., Godley, B.J., Grecian, W.J., Guinet,  
380 C., Hoennner, X., Kienle, S.S., Robinson, P.W., Votier, S.C., Whiting, S., Witt, M.J., Hindell,  
381 M.A., Harcourt, R.G. & McMahon, C.R. (2020) A continuous-time state-space model for  
382 rapid quality control of Argos locations from animal-borne tags. *Movement Ecology*, **8**, 31.
- 383 Jonsen, I.D., McMahon, C.R., Patterson, T.A., Auger-Méthé, M., Harcourt, R., Hindell, M.A.  
384 & Bestley, S. (2019) Movement responses to environment: fast inference of variation among  
385 southern elephant seals with a mixed effects model. *Ecology*, **100**, e02566.
- 386 Jonsen, I., Flemming, J. & Myers, R. (2005) Robust state-space modeling of animal movement  
387 data. *Ecology*, **86**, 2874–2880.
- 388 Kays, R., Crofoot, M.C., Jetz, W. & Wikelski, M. (2015) Terrestrial animal tracking as an  
389 eye on life and planet. *Science*, **348**, aaa2478.
- 390 Kristensen, K., Nielsen, A., Berg, C.W., Skaug, H. & Bell, B.M. (2016) TMB: Automatic  
391 differentiation and Laplace approximation. *Journal of Statistical Software*, **70**, 1–21.
- 392 London, J.M. (2020) *pathroutr: An R Package for (Re-)Routing Paths Around Barriers*.

- 393 McClintock, B.T. & Michelot, T. (2018) momentuHMM: R package for generalized hidden  
394 Markov models of animal movement. *Methods in Ecology and Evolution*, **9**, 1518–1530.
- 395 McMahon, C.R., Roquet, F., Baudel, S., Belbeoch, M., Bestley, S., Blight, C., Boehme, L.,  
396 Carse, F., Costa, D.P., Fedak, M.A., Guinet, C., Harcourt, R., Heslop, E., Hindell, M.A.,  
397 Hoenner, X., Holland, K., Holland, M., Jaine, F.R.A., Jeanniard du Dot, T., Jonsen, I.,  
398 Keates, T.R., Kovacs, K.M., Labrousse, S., Lovell, P., Lydersen, C., March, D., Mazloff,  
399 M., McKinzie, M.K., Muelbert, M.M.C., O'Brien, K., Phillips, L., Portela, E., Pye, J.,  
400 Rintoul, S., Sato, K., Sequeira, A.M.M., Simmons, S.E., Tsontos, V.M., Turpin, V., van  
401 Wijk, E., Vo, D., Wege, M., Whoriskey, F.G., Wilson, K. & Woodward, B. (2021) Animal  
402 borne ocean sensors – anibos – an essential component of the global ocean observing system.  
403 *Frontiers in Marine Science*, **8**, 751840.
- 404 Michelot, T., Langrock, R. & Patterson, T.A. (2016) moveHMM: an R package for the  
405 statistical modelling of animal movement data using hidden Markov models. *Methods in  
406 Ecology and Evolution*, **7**, 1308–1315.
- 407 Michelot, T., Langrock, R., Bestley, S., Jonsen, I.D., Photopoulou, T. & Patterson, T.A.  
408 (2017) Estimation and simulation of foraging trips in land-based marine predators. *Ecology*,  
409 **98**, 1932–1944.
- 410 Patterson, T.A., Thomas, L., Wilcox, C., Ovaskainen, O. & Matthiopoulos, J. (2008) State-  
411 space models of individual animal movement. *Trends in Ecology and Evolution*, **23**, 87–94.
- 412 Pedersen, T.L. (2020) *patchwork: The Composer of Plots*. R package version 1.1.1.
- 413 Phillips, L., Carroll, G., Jonsen, I., Harcourt, R., Brierley, A.S., Wilkins, A. & Cox, M. (2022)  
414 Variability in prey field structure drives inter-annual differences in prey encounter by a  
415 marine predator, the little penguin. *Royal Society Open Science*, **9**, 220028.
- 416 Phillips, S.J., Dudík, M., Elith, J., Graham, C.H., Lehmann, A., Leathwick, J. & Ferrier,

- 417 S. (2009) Sample selection bias and presence-only distribution models: implications for  
418 background and pseudo-absence data. *Ecological Applications*, **19**, 181–197.
- 419 Preisler, H.K., Ager, A.A. & Wisdom, M.J. (2013) Analyzing animal movement patterns  
420 using potential functions. *Ecosphere*, **4**, 1–13.
- 421 R Core Team (2021) *R: A Language and Environment for Statistical Computing*. R Foundation  
422 for Statistical Computing, Vienna, Austria.
- 423 Raymond, B., Lea, M.A., Patterson, T., Andrews-Goff, V., Sharples, R., Charrassin, J.B.,  
424 Cottin, M., Emmerson, L., Gales, N., Gales, R., Goldsworthy, S.D., Harcourt, R., Kato, A.,  
425 Kirkwood, R., Lawton, K., Ropert-Coudert, Y., Southwell, C., van den Hoff, J., Wienecke,  
426 B., Woehler, E.J., Wotherspoon, S. & Hindell, M.A. (2015) Important marine habitat off  
427 east Antarctica revealed by two decades of multi-species predator tracking. *Ecography*, **38**,  
428 121–129.
- 429 Riaz, J., Bestley, S., Wotherspoon, S. & Emmerson, L. (2021) Horizontal-vertical movement  
430 relationships: Adélie penguins forage continuously throughout provisioning trips. *Movement  
431 Ecology*, **9**, 43.
- 432 South, A. (2022) *rnatuearth: World Map Data from Natural Earth*. R package version  
433 0.3.0.
- 434 Thygesen, U.H., Albertsen, C.M., Berg, C.W., Kristensen, K. & Neilsen, A. (2017) Validation  
435 of ecological state space models using the Laplace approximation. *Environmental and  
436 Ecological Statistics*.
- 437 Wickham, H. (2016) *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York.