ExMove: An open-source toolkit for processing and exploring animal tracking data in R

Liam Patrick Langley[1†], Stephen Lang[1], Luke Ozsanlav-Harris[1,2], Alice Trevail[1]

[1]Centre for Ecology and Conservation, University of Exeter, Penryn, Cornwall

[2]Wildfowl and Wetlands Trust, Slimbridge, Gloucester

[†]Corresponding Author

Keywords: Animal movement, Tracking data, Reproducibility, Code sharing, User guide, GPS, geolocator, ARGOS

Abstract

1. Ongoing technological advances have led to a rapid increase in the number, type, and scope of animal tracking studies. In response, many software tools have been developed to analyse animal movement data. These tools generally focus on movement modelling, but the steps required to clean raw data files from different tracking devices have been largely ignored. Such pre-processing steps are often time-consuming and involve a steep learning curve but are crucial for the creation of high quality, standardised, and shareable data. Moreover, decisions made at this early stage can substantially influence subsequent analyses, and in the current age of reproducibility crisis, the transparency of this process is vital.
2. Here we present an open-access, reproducible toolkit written in the programming language R for processing raw data files into a single cleaned data set for analyses and upload to online tracking databases (found here: https://github.com/ExMove/ExMove). The toolkit comprises well documented and flexible code to facilitate data processing and user understanding, both of which can increase user confidence and improve the uptake of sharing open and reproducible code. Additionally, we provide an overview website (found here: https://exmove.github.io/) and a Shiny app to help users visualise tracking data and assist with parameter determination during data cleaning.
3. The toolkit is generalisable to different data formats and device types, uses modern 'tidy coding' practices, and relies on few well-maintained packages. Amongst these, we perform spatial manipulations using the package *sf*, providing a timely resource for movement ecologists following the recent deprecation of the widely-used *sp* package.

4. Overall, by collating all required steps from data collection to archiving on open access databases into a single, robust pipeline, our toolkit provides a valuable resource for anyone conducting animal movement analyses and represents an important step towards increased standardisation and reproducibility in animal movement ecology.

1 Introduction

We live in a golden age of animal-tracking data. Recent technological advancements have reduced the size and cost of telemetry devices and improved their capacities, with more species and individuals tracked over longer periods, allowing researchers to expand the frontiers of ecological research (Bodey et al., 2018; Kays et al., 2015; Nathan et al., 2022; Wilmers et al., 2015). These studies have generated a rapidly expanding volume of data (Gupte et al., 2022; Kays et al., 2022), from an increasing range of device types, with associated formats. This proliferation has complicated the workflow from raw files to a cleaned data set suitable for analysis and sharing, with knock-on impacts for reproducibility. Such pre-analysis steps are universal when working with animal movement data and involve the determination of key filtering parameters to remove data that can influence biological inference from subsequent analysis. Moreover, these steps are time consuming (Dasu & Johnson, 2003), often involving a steep learning curve, and large amounts of time and associated grant funding are lost as unacquainted researchers re-run these steps and, in many cases, attempt to re-invent what should be a standardised and reproducible process.

Numerous software tools have been developed to facilitate animal movement research (Beal et al., 2021; Joo et al., 2020). However, these tools largely focus on evaluating data set suitability (Bjørneraas et al., 2010; Morera-Pujol et al., 2023) and conducting movement analyses of cleaned and/or collated data (Jonsen et al., 2023; Kölzsch et al., 2022). The upstream pre-processing steps required to generate an appropriate cleaned data set have been given scant attention (however see (Gupte et al., 2022). The value of transparent data pre-processing and the accessibility of authentic data sets is important amid the push towards open and reproducible science (Culina et al., 2020; Poisot et al., 2019). Tools to facilitate this move can enhance user's confidence in sharing code (Gomes et al., 2022), and break down barriers to address the reproducibility crisis facing science (Borregaard & Hart, 2016).

There has been progress in developing tools for pre-processing. Gupte et al., (2022) provides an informative framework outlining the key pre-processing steps for high-throughput animal tracking data, alongside an R package, *atlastools*, to implement these steps. MoveApps (Kölzsch et al., 2022) takes an alternative approach to the problem, providing an online user interface where researchers can upload and explore their own tracking data. This provides a tool to rapidly visualise animal tracking data, especially for researchers who lack programming experience. Both approaches do require additional pre-processing of raw data downloads from tags, for example to combine data from multiple individuals, filter data to deployment periods only, and coerce data into specific formats and column names. Collating well-structured data in this way can be complex (Poisot et al., 2019) and poor confidence in code quality to achieve such tasks has contributed to historically low uptake of journal-mandated code archiving (Baker, 2016). Standardised and reproducible pipelines have been highlighted as a key step

towards defensible and robust data science (Borregaard & Hart, 2016; Brunsdon & Comber, 2021), whilst also saving individual researchers time and improving learning (Poisot et al., 2019).

Here, we address this gap by providing a code-based toolkit to process raw tracking files into a clean data set. Establishing a standardised and reproducible pipeline that prepares data for advanced animal movement analyses and archiving in online tracking databases; for example Movebank (Kays et al., 2022) and the Seabird Tracking Database (www.seabirdtracking.org). We have opted for a code-based toolkit shared via versioned GitHub releases rather than an R package, as we want to maximise accessibility and flexibility for a broad user-base. Experienced movement analysts can benefit from rapid data processing, whilst newer users can learn modern R coding skills for processing their tracking data. Our methods involve thorough interrogation of the user-defined parameters for data filtering that can have important downstream consequences for biological inference. By sharing transparent code, we equip users with tools to learn from and adapt for their own analyses. This will make processing tracking data accessible to a wider audience, breaking down financial barriers associated with expensive training courses, while simultaneously flattening the learning curve for ecologists everywhere. The ExMove toolkit aims to equip researchers at all levels with accessible, flexible, and shareable code, and help contribute towards more open-access attitudes in science (Barnes, 2010).

2 Objectives

Given the increase in animal movement research, and current gaps in the availability and accessibility of data pre-processing tools, we have developed a toolkit in the programming language R. This manuscript and the code provided have the following overall objectives:

1. Provide a flexible pipeline to process data from tracking devices, such as Global Positioning System (GPS), light-level geolocators (GLS) and Argos tags, into a clean and processed data set upon which researchers can build their own bespoke animal movement analyses.
2. Streamline the path from data collection to sharing of cleaned data sets in online repositories to facilitate open-access and reproducibility requirements of funders, journals, and institutions.
3. Openly-share a resource for users to learn appropriate R coding tools whilst processing animal telemetry data. This will help to develop user confidence in their code quality and promote code-sharing, leading to improved reproducibility within movement ecology (Gomes et al., 2022; Poisot et al., 2019).
4. Facilitate rapid and efficient parameter determination for data processing by providing an interactive Shiny App, which incorporates a range of useful visualizations to explore data.
5. Maximise stability by using a few well-maintained core R packages, including *here*, *tidyverse* and *sf*. Code that implements these packages is especially relevant given the

widespread adoption of tidy-style coding and the recent deprecation of *sp* (Müller & Bryan, 2020; Pebesma, 2018; Wickham et al., 2019).

3 Workflow

Here, we provide an outline of the code-based workflow from the ExMove toolkit (Figure 1) that enable a user to import raw tracking data and metadata into R and then perform data cleaning, filtering, and processing steps, ready for further analysis. The workflow is flexible and is designed to accommodate anyone undertaking analysis of animal movement data. By following this workflow, users will achieve the following minimum requirements for clean data: date and time formats standardised across time zones, relevant metadata included, no-data values removed (e.g., “NA”, “0,0” in latitude/longitude or other erroneous values), erroneous speeds and locations removed, duplicates removed, locations within pre-/post-deployment window removed, and spatial projections standardised. For a full example of how to use all code within the workflow, please see the user guide (Supplementary Material User guide & https://exmove.github.io). The user guide contains a fully worked example of each step, instructions for accessing resources, and an introduction to the directory structure and data principles that we use to ensure standardisation and reproducibility.

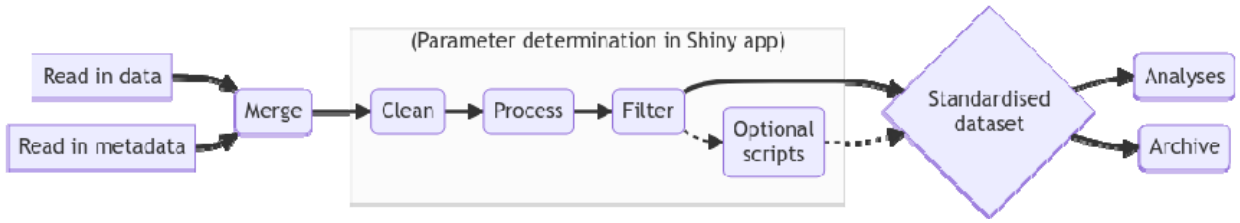
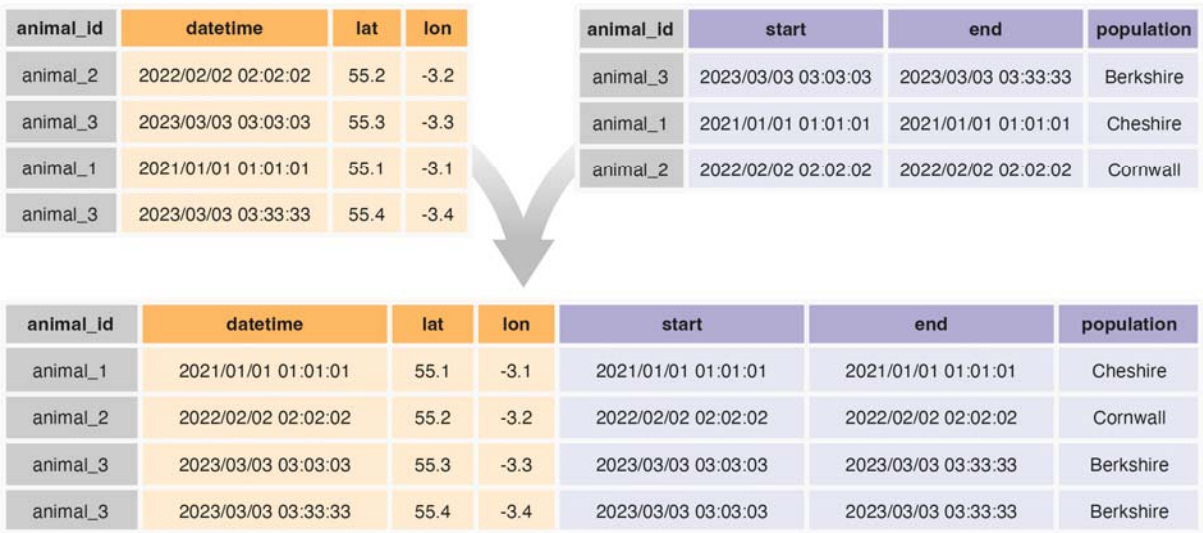


Figure 1. Diagram of workflow used for processing movement data from raw data to archiving (thick line denotes core path of workflow).

3.1 Read in data, Read in metadata, and Merge

During the initial step of the workflow, we read in raw tracking data and metadata files and merge them (user guide sections 1-2). Raw tracking data refers to data downloaded directly from telemetry devices, which fundamentally must contain a time stamp and location data. We also provide alternative code for data that have already been partially processed, for example data downloaded from an existing repository that contains multiple tags and/or individuals. Metadata provide essential technical, biological, and ecological context for the raw data, such as the tag deployment period, sex, age and population. Metadata should contain one row for each tag deployment per individual, with ID numbers corresponding to those in the raw data. The ID numbers are used to match each row of the raw data to the corresponding metadata. Additional code is provided to handle scenarios where a single individual has carried multiple tags in its lifetime, or a single tag has been deployed on multiple individuals. Example data are saved as .csv files (see supplementary files), and guidance is provided for importing data in other text-based formats (e.g., .txt or .tsv). In addition to telemetry data, the code from this

step can also be adapted for biologging studies, to merge and standardise additional sensor data (e.g., water immersion or depth) from multiple individuals.



animal_id	datetime	lat	lon
animal_2	2022/02/02 02:02:02	55.2	-3.2
animal_3	2023/03/03 03:03:03	55.3	-3.3
animal_1	2021/01/01 01:01:01	55.1	-3.1
animal_3	2023/03/03 03:33:33	55.4	-3.4

animal_id	start	end	population
animal_3	2023/03/03 03:03:03	2023/03/03 03:33:33	Berkshire
animal_1	2021/01/01 01:01:01	2021/01/01 01:01:01	Cheshire
animal_2	2022/02/02 02:02:02	2022/02/02 02:02:02	Cornwall

animal_id	datetime	lat	lon	start	end	population
animal_1	2021/01/01 01:01:01	55.1	-3.1	2021/01/01 01:01:01	2021/01/01 01:01:01	Cheshire
animal_2	2022/02/02 02:02:02	55.2	-3.2	2022/02/02 02:02:02	2022/02/02 02:02:02	Cornwall
animal_3	2023/03/03 03:03:03	55.3	-3.3	2023/03/03 03:03:03	2023/03/03 03:33:33	Berkshire
animal_3	2023/03/03 03:33:33	55.4	-3.4	2023/03/03 03:03:03	2023/03/03 03:33:33	Berkshire

Figure 2. Example of combining raw GPS tracking data file (orange) and metadata file (lilac), using the ExMove toolkit, with “animal_id” as the index column. Note that accuracy of latitude and longitude data has been reduced for clarity.

3.2 Cleaning

In the cleaning step, we remove data that are incorrect based on deployment periods and technical tag information (user guide section 3). This includes limiting data to the deployment period (i.e., for loggers programmed to start ahead of deployment or not turned off immediately following retrieval), removing duplicate rows, and removing NA’s or known values that correspond to missing data based on tag manufacturer guidance. Some tags record a

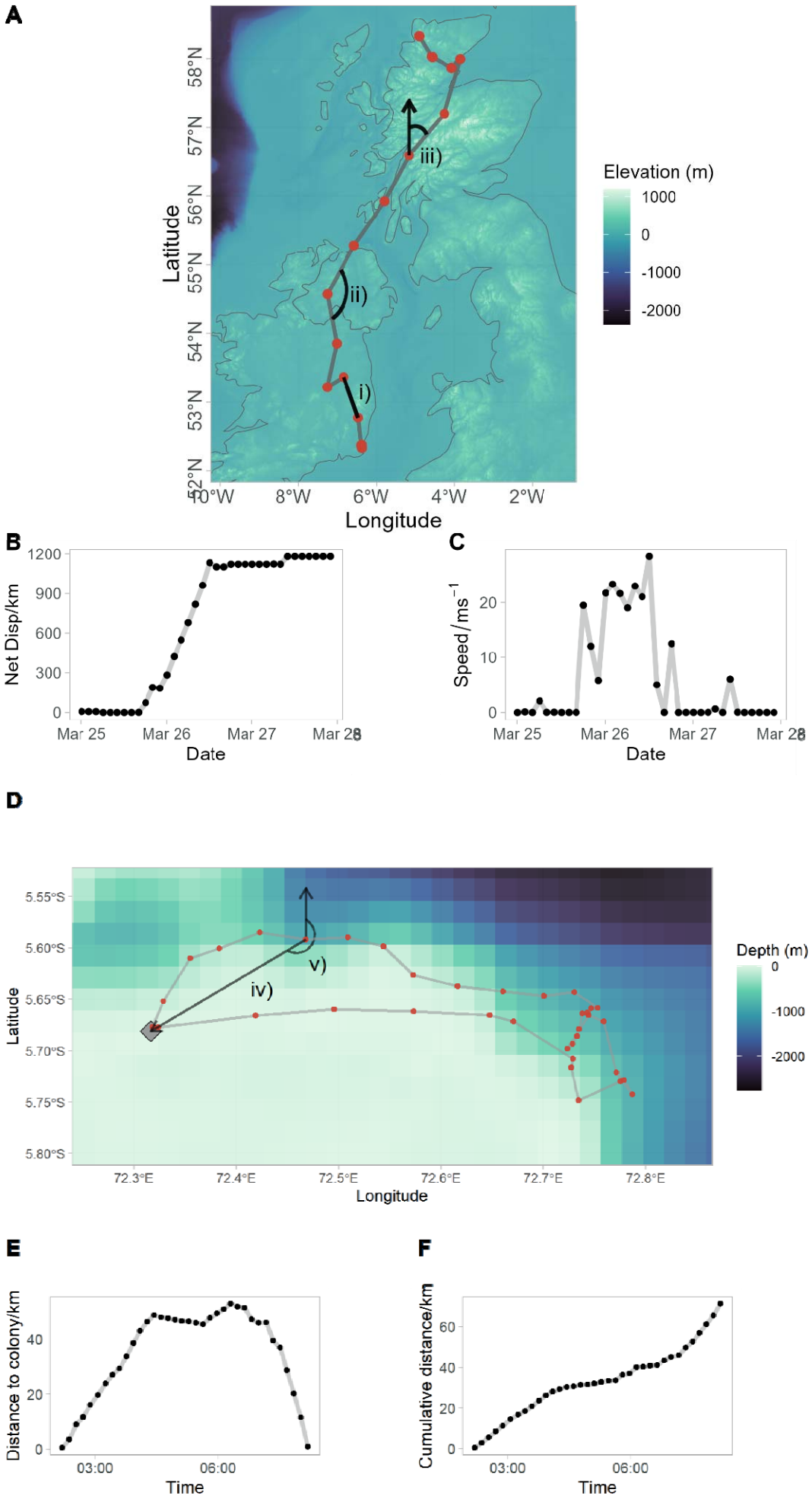
specific ‘no data’ value when sensor information cannot be derived (e.g., “0,0” for latitude and longitude).

3.3 Processing

During the processing step we calculate spatial and temporal metrics to allow downstream analyses and filtering of data based on ecological knowledge and device configuration (user guide section 4). These include metrics pertaining to ecological processes, such as distance travelled (Euclidian distance between successive locations), net displacement (Euclidian distance from first location for each unique ID, e.g., used to quantify migration phenology; (Soriano-Redondo et al., 2020), travel speed (distance travelled dived by time step duration), and turning angle (trajectory of movement at current time step relative to previous trajectory). We also include time step duration (duration between successive locations), which can vary depending on device configuration and performance, but is informative for downstream movement analyses. See Fig. 3 for graphical visualization of these metrics, the code for which is presented in the workflow and incorporated into the shiny app.

Location information from biologging devices is recorded in a specific Coordinate Reference System (CRS). Most often, locations are recorded as degrees of latitude and longitude (CRS: WGS84) which are ‘unprojected’ representations of locations on a sphere, and are generally not recommended for performing distance calculations. ‘Projected’ spatial data (like the format we convert data to in the workflow) are characterised by a CRS that refers to the 2D projection of the earth’s surface used to represent locations. This type of CRS is recommended for calculating

distances and angles and is defined by the user during the processing step. We transform data in the workflow to a global ‘equal area projection’ for spatial calculation, which can be useful for tracking data spanning a large geographic range. However, there is no “correct” CRS and the specific choice may depend on whether users want to preserve distances or angles, and the spatial extent of the data. We provide further guidance and links in section 4 of our FAQs (https://exmove.github.io/FAQs.html).


A
Latitude
Longitude
58°N
57°N
56°N
55°N
54°N
53°N
52°N
10°W
8°W
6°W
4°W
2°W
i)
ii)
iii)
Elevation (m)
1000
0
-1000
-2000
B
Net Disp/km
1200
900
600
300
0
Mar 25
Mar 26
Mar 27
Mar 28
Date
C
Speed / ms⁻¹
20
10
0
Mar 25
Mar 26
Mar 27
Mar 28
Date
D
Latitude
5.55°S
5.60°S
5.65°S
5.70°S
5.75°S
5.80°S
72.3°E
72.4°E
72.5°E
72.6°E
72.7°E
72.8°E
Longitude
iv)
v)
Depth (m)
0
-1000
-2000
E
Distance to colony/km
40
20
0
03:00
06:00
Time
F
Cumulative distance/km
60
40
20
0
03:00
06:00
Time

191

Figure 3. Processed metrics derived from animal tracking data. Plot A depicts a movement track of a Greenland White-fronted Goose with three different metrics labelled: i) step length, ii) turning angle and iii) travel bearing. The net displacement (plot B) and speed (plot C) over time are shown for the movement track in plot A. Plot D depicts a single foraging trip of a Red-footed Booby from a central place (grey filled diamond), illustrating the metrics: iv) distance to central place and v) bearing to central place for a given location. The distance to colony (plot E) and cumulative distance travelled (plot F) are shown for the track in Plot D.

3.4 Filtering

During the filtering step we remove erroneous locations based on three derived variables: 1) speed, 2) net displacement and 3) time since tag deployment (user guide section 6). Each of these variables are described and calculated during the 3.3 processing step (Fig 3). Threshold values are selected by the user to filter the data – if these values are exceeded then the containing rows are removed. The downstream consequences of filtering decision can be explored with the visualizations we provided in our Shiny app (see 4.3). Ecological knowledge from the user is required to help inform thresholds: for instance a species' maximum travel speed, or the time after tag deployment that behaviour is expected to return to normal. We encourage users to explore a range of values and visualise the consequences using maps of movement tracks and histograms of speed and net squared displacement.

3.5 Summarise

Here we provide a summary of the data that the user has retained after the filtering step (user guide section 7). This is an important opportunity to scrutinise the integrity of the data and determine if further filtering is required. The user selects grouping variables for the data — we suggest initially grouping by individual, but broader (e.g., year and population) or narrower (e.g. date) groupings may be required depending on the type of analysis to be undertaken. For each level of the grouping variable, the following summary statistics are provided: number of locations, number of unique tracking days, start and end dates of tracking period, and the sampling frequency locations.

3.6 Post-processing: segmenting, sub-sampling & classifying foraging trips (optional)

Some types of animal movement analyses require additional processing to prepare data (e.g., hidden Markov models (Langrock et al., 2012) and step selection functions (Fieberg et al., 2021)). The toolkit includes code to implement four types of optional processing step: segmenting, sub-sampling, classifying foraging trips and track animation. Segmenting separates tracks into bursts of continuous data with consistent sampling intervals to avoid over-interpolation across large temporal gaps. Tracks are segmented into separate bursts if the time between two locations exceeds a user defined threshold. Then, only bursts that contain a minimum number of locations are kept for further analysis and this threshold is also defined by the user. Sub-sampling aims to standardise the sampling interval across a track by thinning the

data, thus facilitating comparisons among data sets with different original sampling regimes (Fig 4). The user defines a new temporal sampling frequency which should be lower than the average sampling frequency determined at earlier steps (i.e., greater time steps). In cases where increasing sampling frequency, or interpolation, are required, we recommend users to find appropriate methods for their study system and question. For example, data exported from the ExMove toolkit is in an appropriate format for the *aniMotum* package, which uses error radii around telemetry locations to derive most likely movement trajectories (Jonsen et al., 2023).

For central place foragers, we include an optional script to calculate distance and bearing to a central place (typically a breeding colony or roosting site) and classify distinct foraging trips (Trevail et al., 2019). The central place is user-defined and may refer to a known individual nest site, a central colony location for the population, or a shapefile/polygon delineating the spatial boundary of the central place. We provide code to classify and number trips away from the central place using either a distance threshold approach or based on birds moving outside the colony area boundary, and then to classify trips as 'foraging' trips using a time threshold, facilitating the removal of very short trips (e.g., in response to disturbance events). We then calculate standard metrics of foraging trips: total distance travelled, maximum distance from the central place, trip duration, and distal location, to facilitate standardised reporting of central place forager's behavioural ecology. Finally, we include a script to animate tracks allowing users to dynamically visualise movement trajectories of their organisms over time.

All post-processing steps are provided as additional scripts and are designed to follow on from the main code workflow. Scripts can be run iteratively if multiple additional post-processing

steps are required. Each script contains tools to summarise, visualise, and save the post-processed data. The Shiny app can also be used here to aid parameter determination.

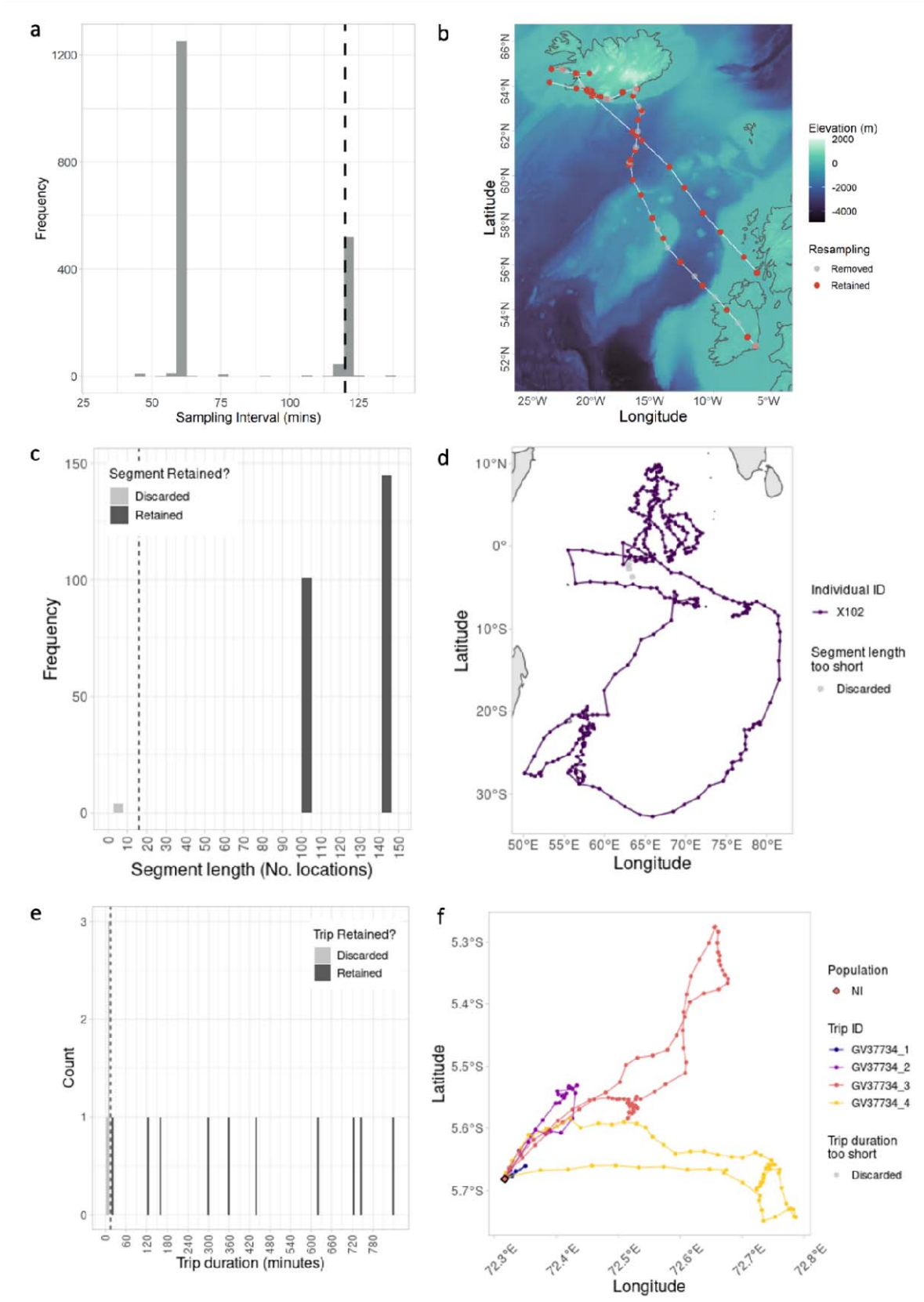


Figure 4. Example of optional post-processing scripts to (a-b) sub-sample data to lower temporal resolution, (c-d) identify segments split by longer temporal intervals, and (e-f) define

central-place foraging trips. (a-b) sampling intervals in Greenland White-fronted Geese data include data collected at one- and two-hour resolutions have been sub-sampled to 2-hour resolution (dashed line in panel a) to standardise data across individuals (red: retained, grey: removed fixes in panel b). (c-d) short segments, containing fewer than 18 tracking locations (dashed line in panel c) have been discarded from the migration track of a Trindade petrel (purple: retained, grey: removed fixes in panel b). (e-f) short trips, less than 15 minutes duration (dashed line in panel e) have been discarded from the central-place foraging trips of a red-footed booby (grey points in panel f are less than 1km from the colony, and so have been obscured by retained trips in distinct colours).

3.7 Standardising output for tracking databases

Open and reproducible science requires data to be presented and archived in a standard, replicable, interoperable way. Online tracking data repositories have achieved this by following the FAIR principles (Wilkinson et al., 2016) and the TRUST principles (Lin et al., 2020). We encourage archiving of tracking data in shared online repositories and as such, we provide code for users to format and save their data, processed within the workflow, in a standardised format that is accepted by two common tracking databases: (1) MoveBank (Kays et al., 2022), and (2) the Seabird Tracking Database.

3.8 Toolkit Continuation

We hope to see widespread adoption of the ExMove toolkit as a starting point for standardised cleaning and processing of animal telemetry data. If users encounter bugs within the toolkit or have suggestions regarding enhancements we encourage them to get in contact with us by raising issues or pull requests via the GitHub. If we make future large-scale changes or add extra functionality we will issue an update via a future GitHub release.

4 Tools and additional practical resources

4.1 User guide

We provide a fully-coded user guide that follows each step of the workflow with worked examples in tidy code. This guide contains example data outputs, plots produced for each step, and detailed annotations to help users adapt the code to their data set. The user guide is available in supplementary material, within the GitHub repository (https://github.com/ExMove/ExMove/tree/main/Documentation), and via the ExMove website (https://exmove.github.io/User_guide.html).

4.2 FAQ's

Some important principles of movement and spatial analyses can be a common source of confusion (for example: understanding data projections). Given ExMove's goal as a learning tool, we provide information to help facilitate better understanding of these core concepts, and links to useful materials made available by others (FAQ page can be found here:

https://github.com/ExMove/ExMove/tree/main/Documentation & https://exmove.github.io/FAQs.html).

4.3 Shiny app

The Shiny app is available online at the following link: https://lukeozsanlav.shinyapps.io/exmove_explorer/. The shiny app follows the same steps as the workflow in sections 3.4 to 3.6 (section 6 onwards of user guide) but they are incorporated into a user friendly user interface. The code for the shiny app is also available in the GitHub repository, alongside the workflow. There is an upload limit so users can either explore the app with examples data sets we provide or a subset of their own data. If the user wants to run the app on their full data set, then we recommend that they clone the repository and then run the app locally on their own machine. The code for doing this is explained in section 6 of the user guide.

The app is designed to be accessible for new R users to visualise and filter data but also provides a convenient interactive platform for more experienced users to determine the downstream effects of user-input parameters on the resulting cleaned data frames. When processing animal tracking data, prior to analysis, users often select rather arbitrary values by which to filter data, i.e., to eliminate erroneous speeds users often remove points creating speeds above a selected threshold. The app allows users to explore the implications of these arbitrary decisions via interactive manipulation of threshold values and subsequent visualizations that include, data summary tables, histograms of the variable of interest and

maps. There are multiple tabs to navigate within the app once users have uploaded their own data, in each there are a series of interactive user input threshold values and a number of plots and tables detailing the consequences of the user selected values. There are tabs for basic data filters (e.g., speed, post-deployment period and date), calculating foraging trips for central placed foragers, sub-sampling location frequency, segmenting movement tracks and finally visualizing the locations in an interactive leaflet map. To facilitate integration with the main workflow there are code chunks output at the bottom of each shiny tab that contain the selected parameter values, which can be copied into the workflow.

4.4 Example data sets

Within the GitHub repo we provide four tested example data sets for users to familiarise themselves with using the toolkit. These data sets are from Trindade Petrels (*Pterodroma arminjoniana*; Figure 5A) breeding on Round Island (Nicoll et al., 2017), immature and adult red-footed boobies (*Sula leucogaster*, Figure 5B & D) breeding in the Chagos Archipelago (Trevail et al 2023) and Greenland white-fronted goose (*Anser albifrons flavirostris*, Figure 5C) migrating from Ireland and Scotland to Greenland (Ozsanlav-Harris et al., 2022). These data sets were selected to illustrate how our toolkit is able to process telemetry data from different tag types, species, life-history stages and spatial movement scales. They also illustrate how running telemetry data through our optional processing scripts can provide additional ecological insight, such as defining foraging trips for central-place foragers (Fig 5D & Fig 6). An example processing

pipeline for central-place foraging red-footed boobies is shown in Figure 6, for which all code is fully presented and described in the user guide.

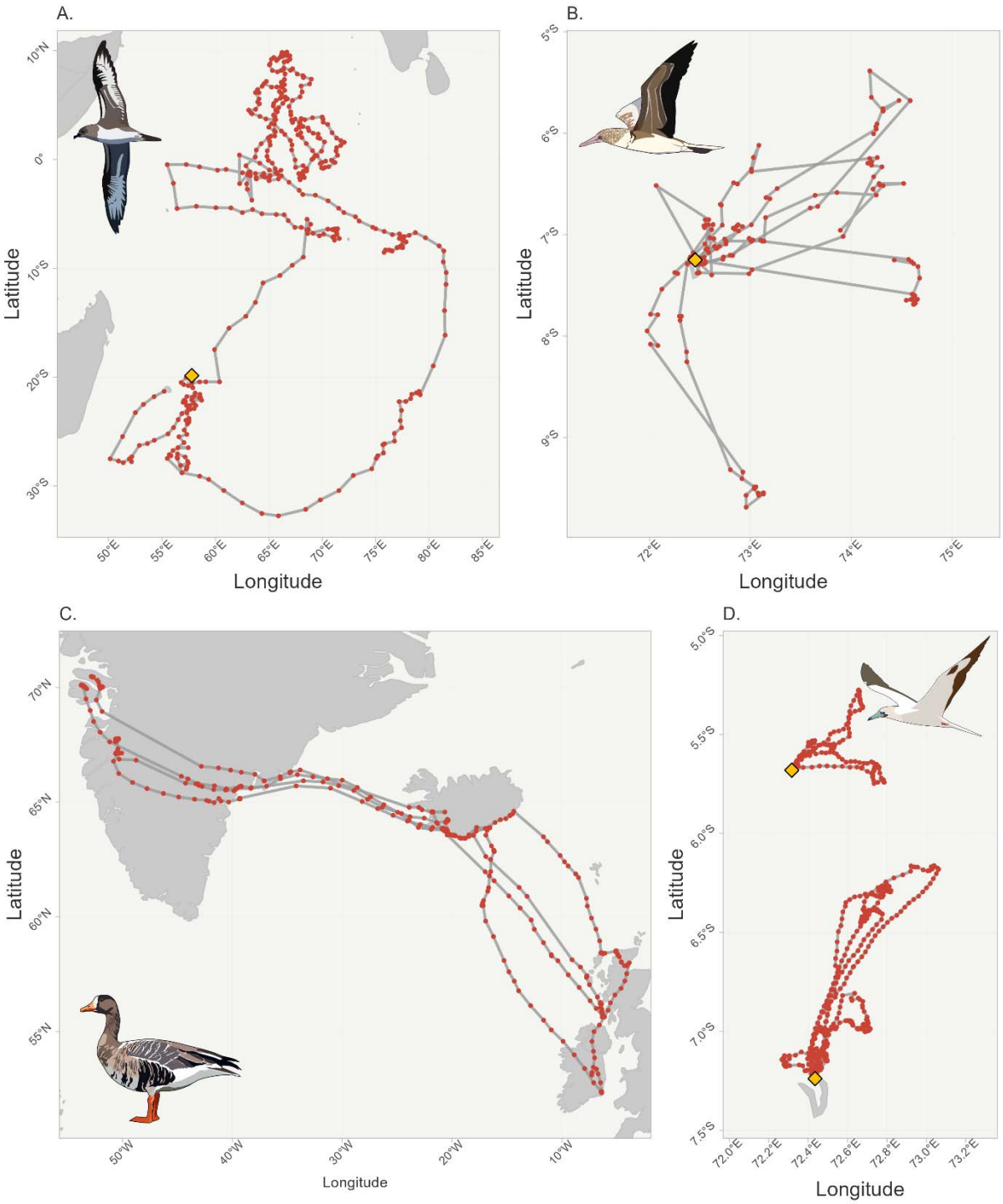


Figure 5. Visualisation of the four example data sets that have been cleaned by the toolkit, with individual telemetry points (red dots) and breeding colony locations (orange diamonds). A: GLS data from a Trindade petrel during its non-breeding migration in 2011. The device was deployed and retrieved from Round Island Nature Reserve (19.85° south; 57.78° east), situated

23 km off the North coast of Mauritius, Indian Ocean. For methods and ethics see Nicoll et al., (2017). B: Argos data from two immature red-footed boobies tracked during 2022 using Microwave Telemetry Solar Argos/GPS PTT loggers (Figure 6B). Data represent foraging movements of immature individuals tagged whilst roosting on East Island in the Chagos Archipelago, Indian Ocean. C: GPS data from four Greenland White-fronted Geese tracked during spring migration in 2018 and 2019. For methods and ethics see Ozsanlav-Harris *et al*. (2022). D: GPS data from three adult red-footed boobies tracked during 2016 and 2018 using igotU loggers (Figure 6D). Data represent central-place foraging trips of breeding adults from two colonies, Diego Garcia and Nelsons Island in the Chagos Archipelago, Indian Ocean. For methods and ethics see Trevail *et al.* (2023).

Central-place foraging red-footed boobies, *Sula sula*

ExMove

Standardised processing pipeline: Workflow.R

Raw tracking data

GV37501_201606_DG_RFB.c
GV37503_201606_DG_RFB.c
GV37734_201807_NI_RFB.csv

1 Import, combine tracking files, parse dates

`df_raw`

2 Metadata

RFB_M

`df_metadataslim`

3 Merge tracking & metadata

```
df_metamerged <- df_raw %>%
  left_join(., df_metadataslim, by=ID_type)
```

4 Clean = remove NAs & keep deployment times

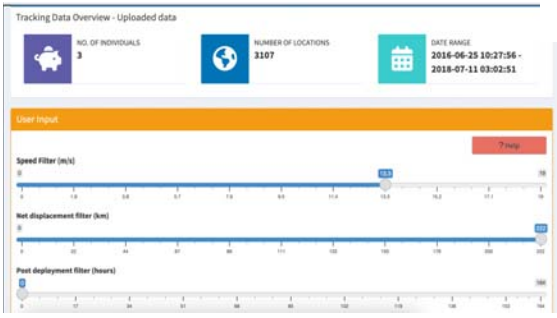
```
df_clean <- df_metamerged %>%
  drop_na(all_of(na_cols)) %>%
  filter(case_when(!is.na(Retrievedatetime) ~
                   Deploydatetime < DateTime &
                   DateTime < Retrievedatetime,
                 .default = Deploydatetime < DateTime))
```

5 Process spatial & temporal metrics

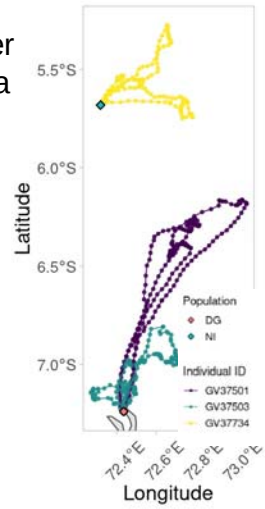
`df_diagnostic`

ID	Y	X	DateTime	Year	Species	Population	Age	BreedStage	dist	difftime	netdisp	speed	turnangle
GV37501	-7.24	72.43	2016-06-28 20:24:59	2016	RFB	DG	Adult	Chick rearing	16.56	425	70.41	0.04	38.42
GV37501	-6.44	72.68	2016-06-28 06:12:56	2016	RFB	DG	Adult	Chick rearing	114.48	309	93972.36	0.37	323.38
GV37503	-7.24	72.43	2016-06-29 01:28:16	2016	RFB	DG	Adult	Chick rearing	17.90	300	51.23	0.06	115.45
GV37503	-7.24	72.43	2016-06-30 14:15:32	2016	RFB	DG	Adult	Chick rearing	2306.19	339	54.25	6.80	124.72
GV37734	-5.68	72.32	2018-07-09 02:21:33	2018	RFB	NI	Adult	Chick rearing	2.21	308	101.29	0.01	49.09
GV37734	-5.68	72.32	2018-07-09 15:29:39	2018	RFB	NI	Adult	Chick rearing	4.56	301	103.07	0.02	348.69

Shiny app: determine filter parameters



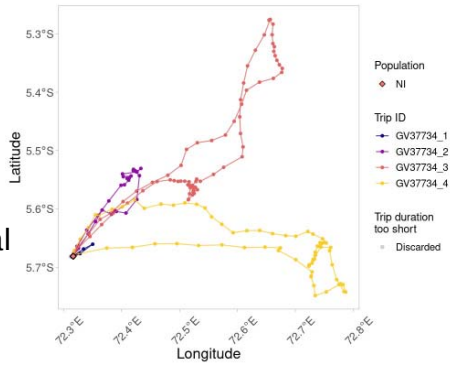
6 Filter data



Define central place trips using optional script

1 Define central place location & trip thresholds

2 Define individual trips



3 Summarise trip metrics

	pulation	N_Trips	Duration	Total_dist	Max_dist
RFB	DG	7	483.56 ± 2.49	148.34 ± 6.43	55.41 ± 4.63
RFB	NI	1	4240.67 ± NA	360.05 ± NA	172.34 ± NA

`df_tripmetrics_summary_pop`

How to report: Tracking data were processed following the ExMove toolkit and Shiny app (Langley et al. 2023) to standardise data, using a speed filter of 20 m/s and a net displacement filter of 300 km, and define central place foraging trips, using a colony radius buffer of 1km, and a minimum trip duration of 15mins

Figure 6. An example of how the ExMove toolkit has been applied to GPS-tracked central-place foraging red-footed boobies, *Sula sula.* The toolkit facilitates import of raw data and metadata, cleaning, processing, and filtering within the `Workflow.R` script. Then identification and summary of central-place behaviour in the `Optional Processing_Central place trips.R` script. The Shiny app aids parameter determination for filtering (speed and net-displacement) and central-place trip definition (distance from colony and minimum trip duration). Results from the tracking study are presented within Trevail et al. (2023).

Acknowledgments

Alice Trevail was funded by the Bertarelli Programme in Marine Science, who also funded collection of adult and immature red-footed booby data. Stephen Lang was funded by the Pew Charitable Trusts. We thank the Wildfowl and Wetlands Trust for providing tracking data from Greenland White-fronted Geese and Larry Griffin for deploying tags. We thank Malcolm Nicoll for valuable comments on an earlier version of the manuscript, and for providing tracking data from a Trindade petrel. The Trindade petrel tracking programme (Round Island, Mauritius) was supported by NERC (Grant NE/H5081500) and Research England, with thanks to Mauritian Wildlife Foundation (MWF) and the National Parks and Conservation Service (NPCS). We also thank Evelyn Alexander, Jacqueline Glencross and Tess Handby for providing valuable feedback on the code, workflow and Shiny app.

Author contributions

All authors contributed equally to the manuscript.

Data availability statement

All data and code presented in this manuscript can be found in the following public GitHub repository release "ExMove 1.0.0" created at the time of publication (GitHub release TBC): https://github.com/ExMove/ExMove. All tracking data sets used are cited appropriately in section 4.4.

References

Baker, M. (2016). 1,500 scientists lift the lid on reproducibility. *Nature*, *533*(7604), 452–454. https://doi.org/10.1038/533452a

Barnes, N. (2010). Publish your computer code: It is good enough. *Nature*, *467*(7317), 753–753. https://doi.org/10.1038/467753a

Beal, M., Oppel, S., Handley, J., Pearmain, E. J., Morera-Pujol, V., Carneiro, A. P. B., Davies, T. E., Phillips, R. A., Taylor, P. R., Miller, M. G. R., Franco, A. M. A., Catry, I., Patrício, A. R., Regalla, A., Staniland, I., Boyd, C., Catry, P., & Dias, M. P. (2021). track2KBA: An R package for identifying important sites for biodiversity from tracking data. *Methods in Ecology and Evolution*, *12*(12), 2372–2378. https://doi.org/10.1111/2041-210X.13713

Bjørneraas, K., Van Moorter, B., Rolandsen, C. M., & Herfindal, I. (2010). Screening Global Positioning System Location Data for Errors Using Animal Movement Characteristics. *Journal of Wildlife Management*, *74*(6), 1361–1366. https://doi.org/10.2193/2009-405

Bodey, T. W., Cleasby, I. R., Bell, F., Parr, N., Schultz, A., Votier, S. C., & Bearhop, S. (2018). A phylogenetically controlled meta-analysis of biologging device effects on birds: Deleterious effects and a call for more standardized reporting of study data. *Methods in Ecology and Evolution*, *9*(4), 946–955. https://doi.org/10.1111/2041-210X.12934

Borregaard, M. K., & Hart, E. M. (2016). Towards a more reproducible ecology. *Ecography*, *39*(4), 349–353. https://doi.org/10.1111/ecog.02493

Brunsdon, C., & Comber, A. (2021). Opening practice: Supporting reproducibility and critical spatial data science. *Journal of Geographical Systems*, *23*(4), 477–496. https://doi.org/10.1007/s10109-020-00334-2

Culina, A., van den Berg, I., Evans, S., & Sánchez-Tójar, A. (2020). Low availability of code in ecology: A call for urgent action. *PLOS Biology*, *18*(7), e3000763. https://doi.org/10.1371/journal.pbio.3000763

Dasu, T., & Johnson, T. (2003). *Exploratory Data Mining and Data Cleaning*. John Wiley & Sons, Inc.

Fieberg, J., Signer, J., Smith, B., & Avgar, T. (2021). A 'How to' guide for interpreting parameters in habitat-selection analyses. *Journal of Animal Ecology*, *90*(5), 1027–1043. https://doi.org/10.1111/1365-2656.13441

Gomes, D. G. E., Pottier, P., Crystal-Ornelas, R., Hudgins, E. J., Foroughirad, V., Sánchez-Reyes, L. L., Turba, R., Martinez, P. A., Moreau, D., Bertram, M. G., Smout, C. A., &

Gaynor, K. M. (2022). Why don't we share data and code? Perceived barriers and benefits to public archiving practices. *Proceedings of the Royal Society B: Biological Sciences*, *289*, 20221113. https://doi.org/10.1098/rspb.2022.1113

Gupte, P. R., Beardsworth, C. E., Spiegel, O., Lourie, E., Toledo, S., Nathan, R., & Bijleveld, A. I. (2022). A guide to pre-processing high-throughput animal tracking data. *Journal of Animal Ecology*, *91*(2), 287–307. https://doi.org/10.1111/1365-2656.13610

Jonsen, I. D., Grecian, W. J., Phillips, L., Carroll, G., McMahon, C., Harcourt, R. G., Hindell, M. A., & Patterson, T. A. (2023). aniMotum, an R package for animal movement data: Rapid quality control, behavioural estimation and simulation. *Methods in Ecology and Evolution*, *14*(3), 806–816. https://doi.org/10.1111/2041-210X.14060

Joo, R., Boone, M. E., Clay, T. A., Patrick, S. C., Clusella-Trullas, S., & Basille, M. (2020). Navigating through the R packages for movement. *Journal of Animal Ecology*, *89*(1), 248–267. https://doi.org/10.1111/1365-2656.13116

Kays, R., Crofoot, M. C., Jetz, W., & Wikelski, M. (2015). Terrestrial animal tracking as an eye on life and planet. *Science*, *348*(6240), aaa2478. https://doi.org/10.1126/science.aaa2478

Kays, R., Davidson, S. C., Berger, M., Bohrer, G., Fiedler, W., Flack, A., Hirt, J., Hahn, C., Gauggel, D., Russell, B., Kölzsch, A., Lohr, A., Partecke, J., Quetting, M., Safi, K., Scharf, A., Schneider, G., Lang, I., Schaeuffelhut, F., … Wikelski, M. (2022). The Movebank system for studying global animal movement and demography. *Methods in Ecology and Evolution*, *13*(2), 419–431. https://doi.org/10.1111/2041-210X.13767

Kölzsch, A., Davidson, S. C., Gauggel, D., Hahn, C., Hirt, J., Kays, R., Lang, I., Lohr, A., Russell, B., Scharf, A. K., Schneider, G., Vinciguerra, C. M., Wikelski, M., & Safi, K. (2022).

MoveApps: A serverless no-code analysis platform for animal tracking data. *Movement Ecology*, *10*(1), 30. https://doi.org/10.1186/s40462-022-00327-4

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*(11), 2336–2342. https://doi.org/10.1890/11-2241.1

Lin, D., Crabtree, J., Dillo, I., Downs, R. R., Edmunds, R., Giaretta, D., De Giusti, M., L'Hours, H., Hugo, W., Jenkyns, R., Khodiyar, V., Martone, M. E., Mokrane, M., Navale, V., Petters, J., Sierman, B., Sokolova, D. V., Stockhause, M., & Westbrook, J. (2020). The TRUST Principles for digital repositories. *Scientific Data*, *7*(1), 144. https://doi.org/10.1038/s41597-020-0486-7

Morera-Pujol, V., Catry, P., Magalhães, M., Péron, C., Reyes-González, J. M., Granadeiro, J. P., Militão, T., Dias, M. P., Oro, D., Dell'Omo, G., Müller, M., Paiva, V. H., Metzger, B., Neves, V., Navarro, J., Karris, G., Xirouchakis, S., Cecere, J. G., Zamora-López, A., … Ramos, R. (2023). Methods to detect spatial biases in tracking studies caused by differential representativeness of individuals, populations and time. *Diversity and Distributions*, *29*(1), 19–38. https://doi.org/10.1111/ddi.13642

Müller, K., & Bryan, J. (2020). *here: A Simpler Way to Find Your Files* (1.0.1) [Computer software]. https://cran.r-project.org/web/packages/here/index.html

Nathan, R., Monk, C. T., Arlinghaus, R., Adam, T., Alós, J., Assaf, M., Baktoft, H., Beardsworth, C. E., Bertram, M. G., Bijleveld, A. I., Brodin, T., Brooks, J. L., Campos-Candela, A., Cooke, S. J., Gjelland, K. Ø., Gupte, P. R., Harel, R., Hellström, G., Jeltsch, F., … Jarić, I. (2022). Big-data approaches lead to an increased understanding of the ecology of

animal movement. *Science*, *375*(6582), eabg1780. https://doi.org/10.1126/science.abg1780

Nicoll, M. A. C., Nevoux, M., Jones, C. G., Ratcliffe, N., Ruhomaun, K., Tatayah, V., & Norris, K. (2017). Contrasting effects of tropical cyclones on the annual survival of a pelagic seabird in the Indian Ocean. *Global Change Biology*, *23*(2), 550–565. https://doi.org/10.1111/gcb.13324

Ozsanlav-Harris, L., Griffin, L. R., Weegman, M. D., Cao, L., Hilton, G. M., & Bearhop, S. (2022). Wearable reproductive trackers: Quantifying a key life history event remotely. *Animal Biotelemetry*, *10*(1), 24. https://doi.org/10.1186/s40317-022-00298-8

Pebesma, E. (2018). Simple Features for R: Standardized Support for Spatial Vector Data. *The R Journal*, *10*(1), 439. https://doi.org/10.32614/RJ-2018-009

Poisot, T., Bruneau, A., Gonzalez, A., Gravel, D., & Peres-Neto, P. (2019). Ecological Data Should Not Be So Hard to Find and Reuse. *Trends in Ecology & Evolution*, *34*(6), 494–496. https://doi.org/10.1016/j.tree.2019.04.005

Soriano-Redondo, A., Acácio, M., Franco, A. M. A., Herlander Martins, B., Moreira, F., Rogerson, K., & Catry, I. (2020). Testing alternative methods for estimation of bird migration phenology from GPS tracking data. *Ibis*, *162*(2), 581–588. https://doi.org/10.1111/ibi.12809

Trevail, A. M., Green, J. A., Sharples, J., Polton, J. A., Miller, P. I., Daunt, F., Owen, E., Bolton, M., Colhoun, K., Newton, S., Robertson, G., & Patrick, S. C. (2019). Environmental heterogeneity decreases reproductive success via effects on foraging behaviour. *Proceedings of the Royal Society B: Biological Sciences*, *286*(1904), 20190795. https://doi.org/10.1098/rspb.2019.0795

Trevail, A., Wood, H., Carr, P., Dunn, R., Nicoll, M., Voiter, S., & Freeman, R. (2023). Multi-colony tracking reveals segregation in foraging range, space use, and timing in a tropical seabird. *Marine Ecology Progress Series.* https://doi.org/10.3354/meps14479

Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L., François, R., Grolemund, G., Hayes, A., Henry, L., Hester, J., Kuhn, M., Pedersen, T., Miller, E., Bache, S., Müller, K., Ooms, J., Robinson, D., Seidel, D., Spinu, V., … Yutani, H. (2019). Welcome to the Tidyverse. *Journal of Open Source Software*, *4*(43), 1686. https://doi.org/10.21105/joss.01686

Wilkinson, M. D., Dumontier, M., Aalbersberg, Ij. J., Appleton, G., Axton, M., Baak, A., Blomberg, N., Boiten, J.-W., Da Silva Santos, L. B., Bourne, P. E., Bouwman, J., Brookes, A. J., Clark, T., Crosas, M., Dillo, I., Dumon, O., Edmunds, S., Evelo, C. T., Finkers, R., … Mons, B. (2016). The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data*, *3*(1), 160018. https://doi.org/10.1038/sdata.2016.18

Wilmers, C. C., Nickel, B., Bryce, C. M., Smith, J. A., Wheat, R. E., & Yovovich, V. (2015). The golden age of bio-logging: How animal-borne sensors are advancing the frontiers of ecology. *Ecology*, *96*(7), 1741–1753. https://doi.org/10.1890/14-1401.1