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

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 analyze 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. Moreover, decisions made at this early stage can substantially influence subsequent analyses.
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 is a codebase that users interact with directly to allow flexibility and understanding of code functionality. Additionally, we provide a Shiny app to help users explore tracking data and provide convenient visualizations to assist with appropriate parameter determination during data cleaning.
3. The toolkit is generalizable to different data formats and device types, uses modern 'tidy coding' practices, and relies on few well-maintained packages. We provide transparent, flexible code to flatten the learning curve associated with animal movement data processing, and produce robust, reproducible datasets.
4. Overall, we envision our resource as a time-saving approach that provides a single reproducible pipeline from data collection to archiving on open access databases, useful for anyone conducting animal movement analyses.

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; 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 and errors. 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 strongly influence the conclusions drawn from subsequent analysis. Moreover, these steps are time consuming, 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 standardized 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 dataset suitability (Bjørneraas et al., 2010; Morera-Pujol et al., 2023) and conducting movement analyses of cleaned and/or collated data (e.g., Jonsen et al., 2023; Kölzsch et al., 2022), and the upstream filtering and processing steps required to generate an appropriate 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 being realised

amid the push towards open and reproducible science. Software tools to facilitate this move are urgently required.

There has been progress in developing tools for pre-processing. Gupte *et al.* (2022) provides an informative framework outlining the key pre-processing steps involved in processing high-throughput animal tracking data, alongside a pipeline to implement these steps. However, the associated atlastools R package wraps these steps up into opaque functions, which may be challenging for less-experienced analysts to de-bug. Additionally, bespoke functions, whilst convenient in many instances, essentially constitute a black box and may provide an impediment to learning. Over-reliance on existing functions may hinder researchers in their effort to develop a foundational understanding of their animal tracking data and the steps required to process it. 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 visualize animal tracking data, especially for researchers who lack programming experience, however the background code is not immediately apparent to users, which again limits the scope for learning. Moreover, MoveApps does not allow upload of raw data from tags, requiring pre-processing into specific formats. Existing approaches demonstrate the need for resources that integrate data processing goals with user learning and development, alongside trends towards open and reproducible science (Brunsdon & Comber, 2021).

Here, we fill this gap by providing a toolkit of foundational code to process raw tracking files into clean data sets, thereby establishing a reproducible starting point for animal movement analysis (found here: https://github.com/ExMove/ExMove). Moreover, we format these data

sets for upload to two widely used online tracking databases; Movebank (Kays et al., 2022) and the Seabird Tracking Database (www.seabirdtracking.org). In doing so we provide a single pathway for creating a standardised dataset for analyses and archiving. Our methods focus on accessible and robust data processing and thorough interrogation of user-defined parameters, which 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. Limited learning opportunities and confidence in code quality have contributed to historically low uptake of journal-mandated code archiving (Baker, 2016), and this will further contribute towards more open-access attitudes in science (Barnes, 2010; Potts & Börger, 2023).

2 Objectives

Given the increase in animal movement research, and current gaps in availability and accessibility of data 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 GPS, GLS and Argos, into a foundational clean data set upon which researchers can build their own bespoke animal movement analyses.

2. Streamline the path from data collection to sharing of cleaned datasets in online repositories to facilitate open-access and reproducibility requirements of funders, journals, and institutions.
3. Increase accessibility to new R users by providing an interactive shiny app that is still useful for more advanced R users by assisting with appropriate parameter determination and visualizations for data cleaning.
4. Generate a learning tool for users to develop skills in animal movement analysis. We provide code organized into annotated code chunks which prioritize user understanding rather than wrapping code up in opaque functions.
5. Maximize stability by using a few well-maintained core R packages, including *here*, *tidyverse* and *sf*. This is timely given the widespread adoption of tidy-style coding and deprecation of *sp* (Müller & Bryan, 2020; Pebesma, 2018; Wickham et al., 2019).
6. Provide open source code, the initial steps of which can be adapted for biologging studies, such as merging and standardizing additional sensor data (e.g., immersion/TDR) from multiple individuals.

By following this workflow, the user will achieve the following minimum requirements for clean data which we define as:

- Date and time formats standardised across time zones
- Rows with corrupted timestamps removed
- Relevant metadata included
- No-data values removed (e.g., “NA”, “0,0” in lat/longs or other erroneous sensor values)

- Duplicates removed
- Undeployed locations removed (e.g., before tag attachment)
- Spatial projections standardised
- Erroneous speeds and locations removed
- Locations at start of tracking period when behaviour is likely to be influenced by capture removed

3 Preliminary Setup

Producing open-access and reproducible analyses is now a common goal for scientific research (Culina et al., 2020; McKiernan et al., 2016), and there are various ways to version and publicly share code (such as git & GitHub). Here, we outline the core best practices that we have followed to help to ensure that the code itself is reliable and can be reproduced by others.

3.1 Preparing your machine and R libraries

In order to make use of the included materials, users must have downloaded R and the RStudio IDE (Ihaka & Gentleman, 1996). Our code is compatible with R versions 4.0.3 - 4.2.1. In general, it is recommended that users are not more than two sub-versions behind the most recent release of R. We make use of the most recent releases of core packages in the *tidyverse* (Wickham et al., 2019), which is composed of *ggplot2* (v3.4.0), *dplyr* (v1.0.10), *tidyr* (v1.2.1), *readr* (v2.1.2), *purrr* (v1.0.1), *tibble* (v3.1.8), *stringr* (v1.5.0) and *forcats* (v0.5.1), as well as *lubridate*

(v1.8.0). We also use *sf* (v1.0-9) for spatial analyses, and the *here* package (v1.0.1) for making reproducible file paths (Müller & Bryan, 2020; Pebesma, 2018).

3.2 Accessing the GitHub resources

The ExMove toolkit can be found in the following GitHub repository: https://github.com/ExMove/ExMove. It can be utilised in a variety of ways, accessible to all levels of experience and desired user outcomes. Users can either: i) 'clone' an updatable version of the ExMove repository via git, ii) download a static copy of the entire repository by navigating to <> Code > Download ZIP, or iii) view and download individual scripts as required. To maximise reproducibility and future functionality, our recommended approach would be to set up R/RStudio with git version control, set up and connect a personal GitHub account, then clone the ExMove repository. Additional information and links for on getting set up with GitHub and RStudio can be found in the ExMove repository readme.

3.3 Directory structure

Directory structure relates to how folders and sub-folders are arranged. We recommend following the directory structure outlined by Figueiredo *et al.* (2022), which suggests that all data required for a project should be inside a single directory, with separate sub-folders for raw data, processed data, and outputs (such as tables and figures). Folders are best named in snakecase format (e.g., "raw_data") to avoid issues with encoding spaces in filepaths (Trisovic et al., 2022). For R Studio users, this folder structure is enhanced with the use of R Studio projects, whose location denotes the 'top' of the directory allowing for shorter, reproducible

filepaths. We also use the here package to maximise reproducibility between operating systems that use different file path delimiters (Müller & Bryan, 2020). When used in combination, these best practices ensure that when sharing or archiving a single project folder, other users have all the data required to recreate your analyses, and results will be reproducible regardless of their machine.

3.4 Data structure

Raw animal tracking data should be saved into a "raw_data" folder in the exact format they are downloaded or decompressed from devices (i.e., completely unaltered). These files should be considered read-only and must not be opened with other software (such as Excel) which can cause problems in changing data formats. Data from different devices will usually arrive in separate files and as such there should be one file per individual (though some cases and data sources may mean that data are combined in some way). We recommend that each individual file should include a unique identifier in the filename. Note that raw data from some GLS devices is provided as light levels and pre-processing may be required to obtain spatial co-ordinates before the data is processed using this workflow (Lisovski & Hahn, 2012; Sumner et al., 2009, p. 2; Thiebot & Pinaud, 2010). Metadata files should also be saved in raw_data as a single file per dataset containing all additional data for each tag deployment, for example individual ID, tag ID, tagging location/time or location of central place if applicable.

The tidyverse is an increasingly used approach for coding in R, with a collection of core packages that use consistent grammar and data structures that allows user to quickly create

analyses that can be easily understood by others (Wickham et al., 2019). To follow tidyverse coding practices, data should adhere to three core rules: i) every variable must have its own column, ii) every observation must have its own row, and iii) every value must have its own cell (Wickham, 2014). Tracking data inherently follow this tidy structure, with rows as observations, with at minimum require a column for time, plus additional columns for variables associated with each time point (e.g., latitude and longitude, depth, temp etc.). Finally, focusing on using core, well-maintained packages (specifically *here*, *tidyverse* and *sf*), helps to maximise code longevity.

4 Workflow

Here, we provide an outline of the data workflow structure and methods (Figure 1) that enable a user to import raw tracking data and metadata, perform basic to advanced cleaning, filtering and processing steps, ready for further scientific analysis. We will introduce the required data structure and outline the flexibility and utility of the workflow for anyone undertaking analysis of animal movement data. 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).

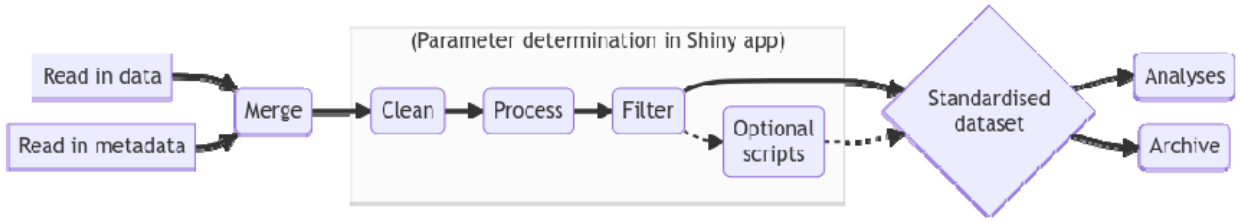


Figure 1: Diagram of workflow used for processing movement data from raw data to archiving (thick line denotes core path of workflow)

4.1 Read in data, Read in metadata, and Merge

The workflow starts with reading in raw tracking data and metadata files. 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., such as .txt or

.tsv).

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 mutating join of raw GPS tracking data file (orange) and metadata file (lilac) using left_join, with “animal_id” as the index column. Note that accuracy of latitude and longitude data has been reduced for clarity.

4.2 Cleaning

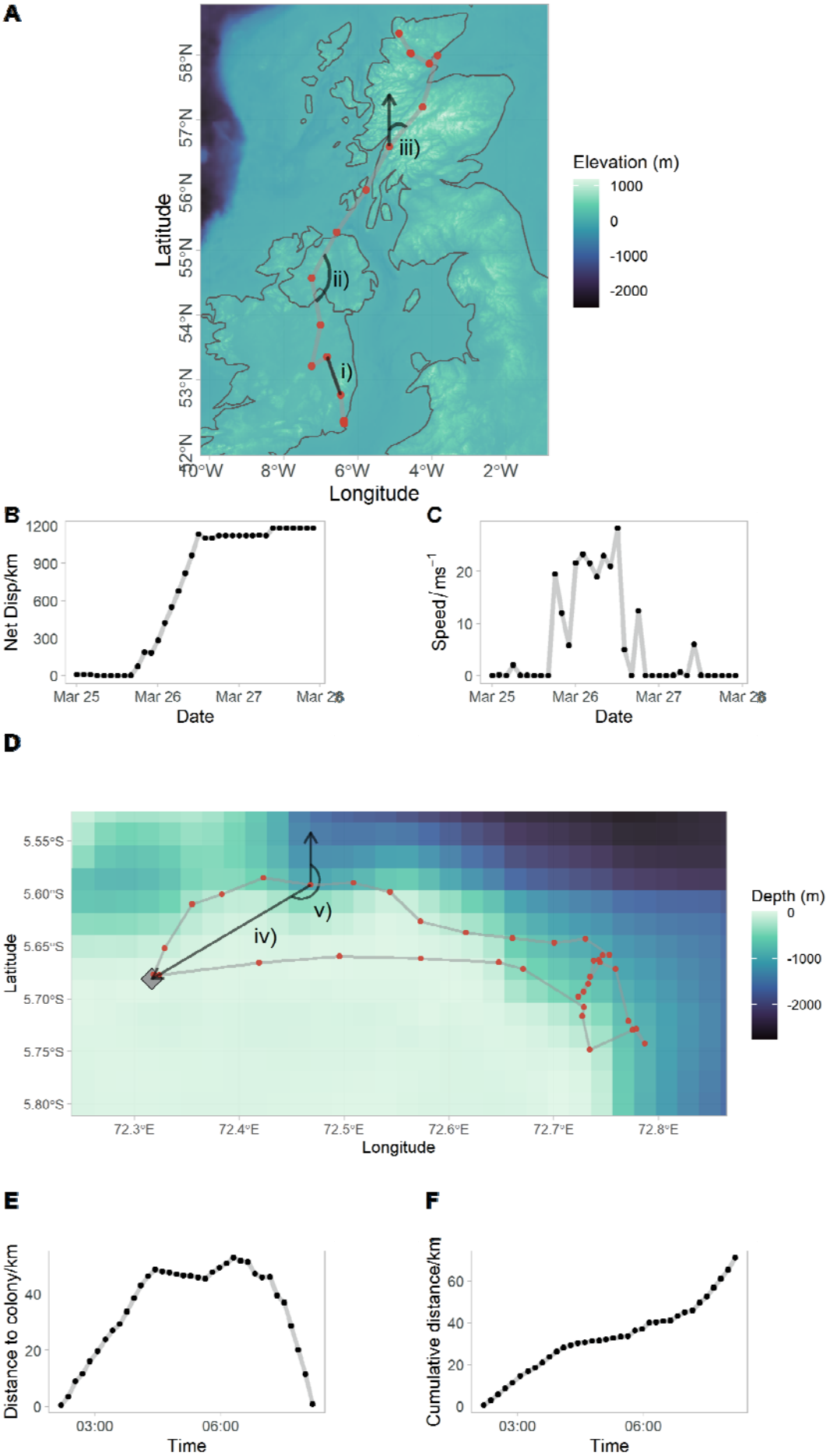
The cleaning step removes data that are incorrect based on deployment periods and technical tag information. This includes limiting data to the deployment period (i.e., for loggers programmed to start ahead of deployment or not turned off immediately following retrieval), removing duplicate rows, and 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).

4.3 Processing

During the processing step we calculate spatial and temporal metrics to allow downstream analyses and filtering of data based on ecological knowledge. These include metrics pertaining to ecological processes, such as distance traveled (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), time step duration (duration between successive locations), travel speed (distance travelled dived by time step duration) and turning angle (trajectory of movement at current time step relative to previous trajectory) and travel bearing (trajectory of movement relative to true north) (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 particular coordinate reference system (CRS) that the user should know. Often locations are recorded as latitudes and longitudes (WGS84 coordinate system) which are un-projected and represent locations on a sphere and are generally not recommend for calculating distances. Projected spatial data (like the format we convert data to in the workflow) possess a coordinate reference system (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. During the processing step the user must define the CRS that they wish to use for calculating distances and angles. We transform data in the workflow to a global 'equal area projection' for spatial calculation, which can be useful for tracking data spans a large geographic range. However there is no "correct" CRS and the specific choice may depend on whether users want to preserve differences 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
58°N
57°N
56°N
55°N
54°N
53°N
52°N
10°W
8°W
6°W
4°W
2°W
Longitude
i)
ii)
iii)
Elevation (m)
1000
0
-1000
-2000
B
Net Disp/km
1200
900
600
300
0
Mar 25
Mar 26
Mar 27
Date
C
Speed / ms⁻¹
20
10
0
Mar 25
Mar 26
Mar 27
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

249

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 is shown for the movement track plot A. Plot D depicts a single foraging trip of a Red-footed Booby from a central place (grey filled diamond). Two metrics are labelled here: iv) distance to central place and v) bearing to central place. The distance to colony (plot E) and cumulative distance traveled (plot F) are shown for the track in Plot D.

4.4 Filter

The filter step allows users to remove erroneous locations based on three derived variables: 1) speed, 2) net displacement and 3) time since tag deployment. Each of these variables are described and calculated during the 4.3 processing step (Fig 3). Threshold values are selected by the user to filter the data, if these values are exceeded then the particular rows are removed. The downstream consequences of these decision can be explored with the visualizations we provide and within the shiny app. Ecological knowledge from the user is required to help inform the thresholds selected, for instance a species' maximum travel speed or the time after tag deployment that behavior returns to normal. We encourage users to explore a range of values and visualize the consequences using maps of movement tracks and histograms of speed and net squared displacement.

4.5 Summarise

This step provides a full summary of the data that the user has retained after the filtering step. This is an important opportunity to scrutinize 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.

4.6 Post-processing: re-sampling, segmenting, & central place foragers (optional)

Some types of animal movement analyses require additional processing to prepare data (e.g. hidden Markov models and step selection functions). The toolkit includes code to conduct three-post processing steps; segmenting, sub-sampling and classifying foraging trips. Segmenting separates tracks into bursts of continuous data with consistent sampling intervals to avoid over-interpolation across large gaps. Tracks are segmented into separate bursts if the time between two locations exceeds a user defined threshold. Then only burst that contain a minimum number of locations are kept for further analysis and this threshold is also defined by the user. Sub-sampling down samples tracking data to facilitate comparisons among datasets with different sampling regimes (Fig 4). The user defines a new sampling frequency which should be greater than the average sampling frequency determined at earlier steps.

For central place foragers, we include an optional script to calculate distance and bearing to a central place (typically a breeding colony) and classify distinct foraging trips (e.g., Trevail et al., 2019). The central place is user defined and may refer to a known nest site, a central colony location for the population, or a shape file delineating the spatial boundary of the central place. We provide code to classify and number trips away from the central place by using distance or a shape file, 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 standardized reporting of central place forager's behavioral ecology.

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.

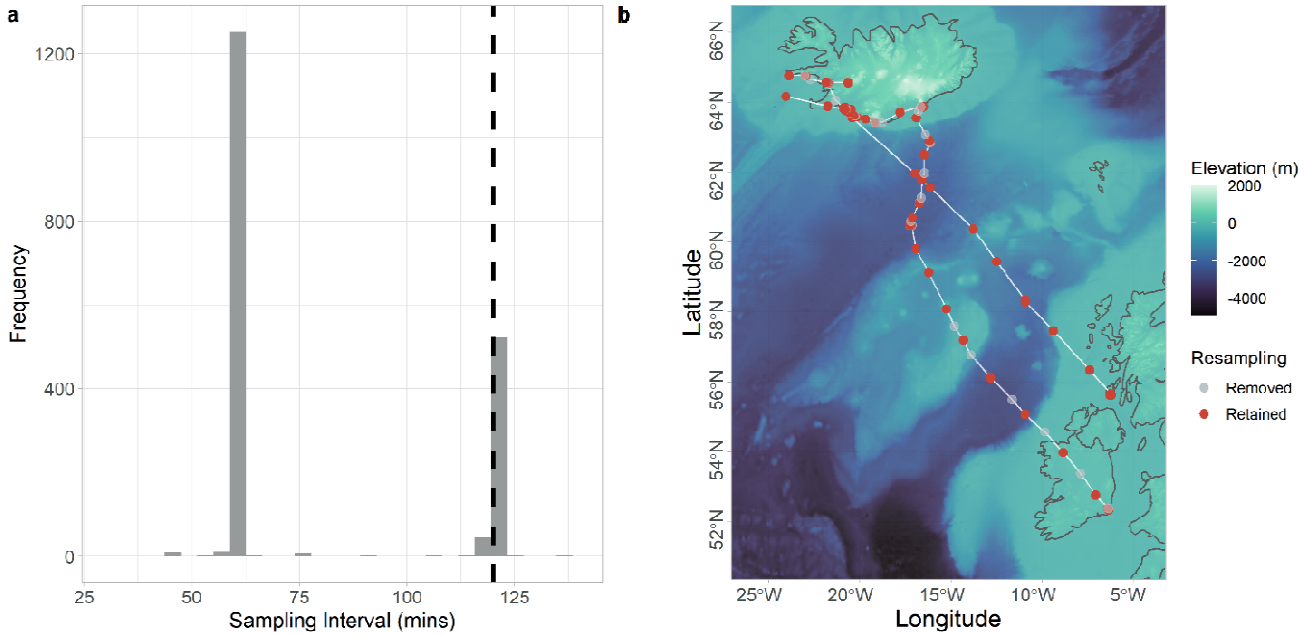


Figure 4. Example of sub-sampling GPS tracking data using an optional post-processing step. The data used are from two Greenland White-fronted Geese migrating from the United Kingdom to Iceland. Panel a depicts a histogram of sampling intervals present in the data, clearly there is a mixture of data collected at one- and two-hour resolution. Therefore, to make sampling intervals consistent the data were sub-sampled to 2 hours (dashed vertical line). In panel b the output of this sub-sampling is visualized and retained (red) and removed (grey) GPS fixes are shown. The individual starting from a more southerly location was originally collected at a 1-hour resolution hence roughly every other fix is removed.

4.7 Standardising output for tracking databases

Many users may need to archive their data on an online database, which often require data to be in a specific format before files can be uploaded (Kays et al., 2022). As such, we provide code for users to format and save their data, processed within the workflow, in the required format of two common tracking databases: (1) MoveBank, and (2) the Seabird Tracking Database.

5 Tools and additional practical resources

5.1 User guide

We provide a fully-coded user guide that follows each step of the workflow, with worked examples in baseR and tidy code. This guide contains example outputs of data frames and plots

produced for each step. The user guide is available as a supplementary material, within the GitHub repository (https://github.com/ExMove/ExMove/tree/main/Documentation), and via the ExMove website (https://exmove.github.io/userguide).

5.2 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).

5.3 Shiny app

The shiny app is available online at the following link: https://lukeozsanlav.shinyapps.io/exmove_explorer/. The shiny app uses the same code as the main workflow from section 6 onwards (including optional processing scripts), incorporated into a user friendly UI. 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 visualize data and data filtering 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 (Fig 5). 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.

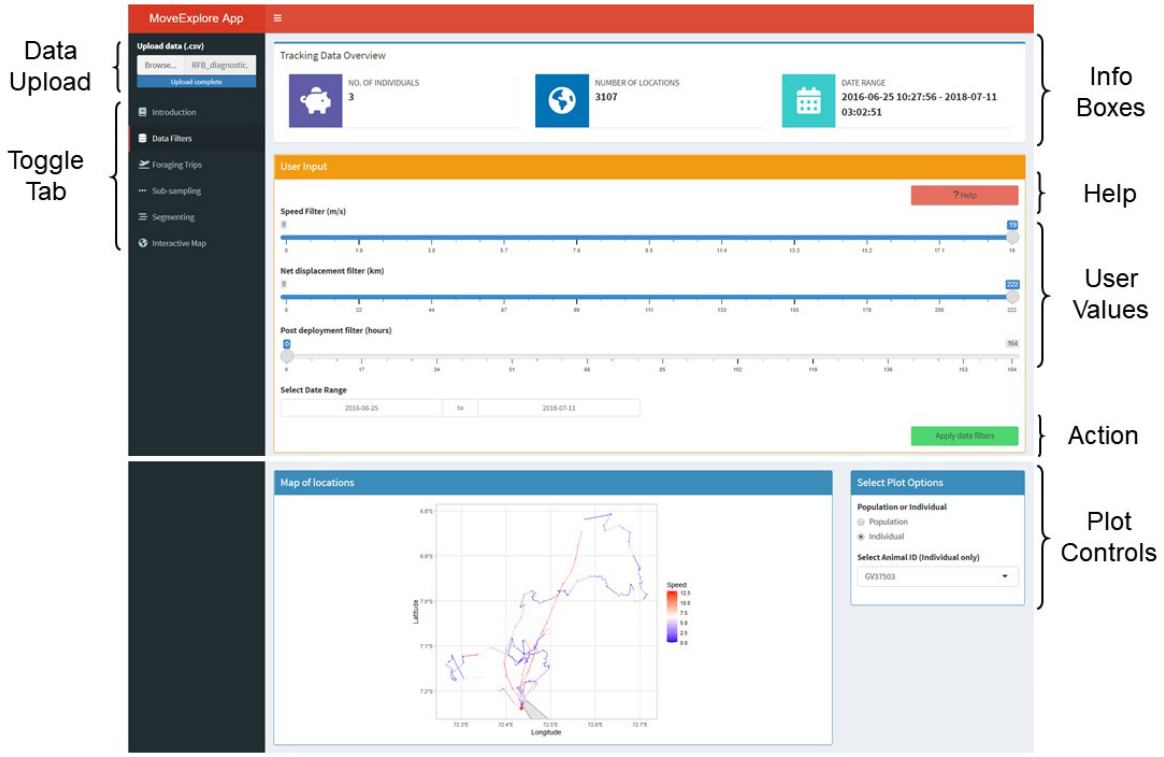


Figure 5. Shiny app to explore and visualise tracking data and test the influence of parameter selection on filtering and post-processing. The data filters tab is depicted here, and each labelled element is described in more detail here. Data upload: search the users system for a file to upload to the app, this should be the equivalent of df_diagnostic. Toggle tab: toggle between different tabs in the app, for each tab the user input box and outputs will change. Info Boxes: summary information of the tracking data that is calculated automatically for the tracking uploaded by the user. User Values: Values that the user can alter that specify the parameters in filtering or post-processing steps. These maybe be sliders, buttons, dropdown menus or calendar date selectors Help; Click to access a pop-up help page that guides users through each tab. Action; press this button to apply the current user value selection to the data. Plot Controls: change whether all or a single individuals is depicted in a map. The specific individual visualised can be chosen using a dropdown menu

5.5 Example datasets

Within the GitHub repo we provide four road-tested example datasets for users to familiarise themselves with running the pipeline. These datasets are from Trindade Petrels (*Pterodroma arminjoniana*; Figure 6A) breeding on Round Island (Nicoll et al., 2017), immature and adult red-footed boobies (*Sula leucogaster,* Figure 6B & D) breeding in the Chagos Archipelago (Carr et al., 2022) and Greenland white-fronted goose (*Anser albifrons flavirostris*, Figure 6C) migrating from Ireland and Scotland to Greenland (Ozsanlav-Harris et al., 2022).

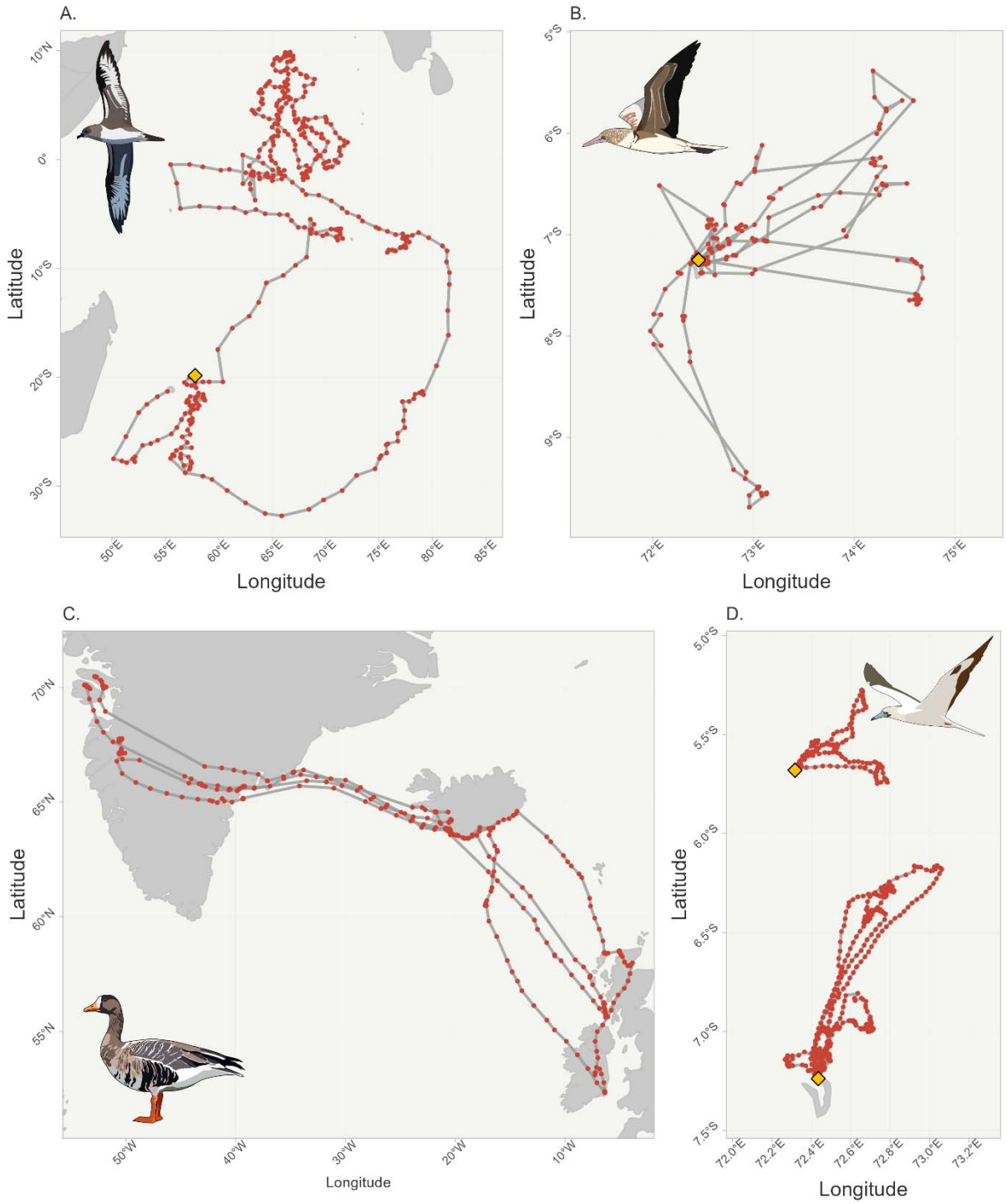


Figure 6. Visualisation of the four example datasets provided as part of the toolkit. Individual locations are shown in red and colony locations shown in orange diamonds for seabirds. A: GLS data from a Trindade petrel tracked during its non-breeding migration in 2011. The device was deployed and retrieved when the individual was resting on Round Island Nature Reserve

(19.85° south; 57.78° east), situated 23 km off the North coast of Mauritius, Indian Ocean. Full methods and ethics are detailed in 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 movements of individuals that have not yet recruited into a breeding population, tagged whilst roosting on East Island in the Chagos Archipelago, central Indian Ocean. C: GPS data from four Greenland White-fronted Geese tracked during spring 2018 and 2019. These data cover the spring migration from Ireland and Scotland to Greenland with a stopover in Iceland. Full methods and ethics are detailed in 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, central Indian Ocean. Full methods and ethics are detailed in Carr *et al.* (2022).

Acknowledgments

Alice Trevail was funded by the Bertarelli Programme in Marine Science, who also funded collection of adult and immature red-footed booby data. 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: https://github.com/ExMove/ExMove. All tracking data sets used are cited appropriately in section 5.5.

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., Moorter, B., Rolandsen, C. M., & Herfindal, I. (2010). Screening Global Positioning System Location Data for Errors Using Animal Movement Characteristics. *The Journal of Wildlife Management*, *74*(6), 1361–1366. https://doi.org/10.1111/j.1937-2817.2010.tb01258.x

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.

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

Carr, P., Trevail, A. M., Koldewey, H. J., Sherley, R. B., Wilkinson, T., Wood, H., & Votier, S. C. (2022). Marine Important Bird and Biodiversity Areas in the Chagos Archipelago. *Bird Conservation International*, 1–8. https://doi.org/10.1017/S0959270922000247

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

Figueiredo, L., Scherer, C., & Cabral, J. S. (2022). A simple kit to use computational notebooks for more openness, reproducibility, and productivity in research. *PLOS Computational Biology*, *18*(9), e1010356. https://doi.org/10.1371/journal.pcbi.1010356

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

Ihaka, R., & Gentleman, R. (1996). R: A Language for Data Analysis and Graphics. *Journal of Computational and Graphical Statistics*, *5*(3), 299–314. https://doi.org/10.1080/10618600.1996.10474713

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–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

Lisovski, S., & Hahn, S. (2012). GeoLight—Processing and analysing light-based geolocator data in R. *Methods in Ecology and Evolution*, *3*(6), 1055–1059. https://doi.org/10.1111/j.2041-210X.2012.00248.x

McKiernan, E. C., Bourne, P. E., Brown, C. T., Buck, S., Kenall, A., Lin, J., McDougall, D., Nosek, B. A., Ram, K., Soderberg, C. K., Spies, J. R., Thaney, K., Updegrove, A., Woo, K. H., & Yarkoni, T. (2016). How open science helps researchers succeed. *ELife*, *5*, e16800. https://doi.org/10.7554/eLife.16800

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.

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

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

Potts, J. R., & Börger, L. (2023). How to scale up from animal movement decisions to spatiotemporal patterns: An approach via step selection. *Journal of Animal Ecology*, *92*(1), 16–29. https://doi.org/10.1111/1365-2656.13832

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

Sumner, M. D., Wotherspoon, S. J., & Hindell, M. A. (2009). Bayesian Estimation of Animal Movement from Archival and Satellite Tags. *PLoS ONE*, *4*(10). https://doi.org/10.1371/journal.pone.0007324

Thiebot, J., & Pinaud, D. (2010). Quantitative method to estimate species habitat use from light-based geolocation data. *Endangered Species Research*, *10*(1), 341–353. https://doi.org/10.3354/esr00261

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

Trisovic, A., Lau, M. K., Pasquier, T., & Crosas, M. (2022). A large-scale study on research code quality and execution. *Scientific Data*, *9*(1), 60. https://doi.org/10.1038/s41597-022-01143-6

Wickham, H. (2014). Tidy Data. *Journal of Statistical Software*, *59*(10). https://doi.org/10.18637/jss.v059.i10

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

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