Design Plan

Daniel Petterson

Resolving Issues With Tracking Seabird Species Using Light-based Sensors

The development of global location sensors (GLS) has brought an alternative way to track species for which a satellite based tracker would have been unsuitable (Wilson 1992). GLS tags can record light intensity, conductivity and in some cases, temperature data. This package aims to remedy issues present in the collection of packages used to read in, process and visualise data from light level loggers/global location sensors (GLS).

Background

Using the features of the light data it is possible to estimate the timing of sunrise and sunset - herein referred to as twilight events - using a threshold method as described by Ekstrom et al. (2004). Light intensities above or below a set threshold signal the occurrence of a twilight event Hill and Braun (2001). As the Earth moves around the Sun, the length of the day (defined as the time between sunrise and sunset) changes. The extent to which it changes depends on the latitude at which the recording was made. Similarly, longitude can be estimated based on the time at which solar noon and midnight - the midpoint between each set of twilight events - occurs. The threshold method is shown to be robust but has notable difficulties near the equator Ekstrom et al. (2004) and around an equinox where there can be little to no variation in day length across a large section of the earth.

Using the `prob_algorithm` function in probGLS (Merkel et al. 2016) it is possible to estimate two locations per full day cycle. The function takes in a multitude of arguments including the known starting point of a bird, parameters for a distribution concerning their speed and that of the local ocean currents as well as the twilight times to generate a cloud of n points for the possible location at the next midday/midnight. The median value in the cloud is then taken as the most likely location. Using information on a species global range we can implement a bounding box to restrict estimated points to a given area. As previously noted, there may be issues with estimating latitude near each equinox as it is dependent on the estimated day length. The probGLS algorithm addresses this by randomly sampling from a uniform distribution between the maximum and minimum latitude of the bounding box (ignoring the speed distribution sampling) and adding them to each computed longitude estimate. This method works well when tracking creatures with a small range or that are known to never cross the equator but for seabirds that migrate across hemispheres it can result in one location estimate being on the other side of the planet from the previous estimate and so estimates during these periods can be misleading. For species with a large longitudinal range in one hemisphere but not the other this may even result in location estimates in areas that the species has never been observed before.

Functions to Import and Transform Data

Light-level loggers for use in GLS tracking are produced by a handful of manufacturers each with their own file types and output format. In addition to this the technical capabilities differ based on device type with some devices also capable of recording temperature and whether the device is submerged. Existing functions from other packages can fail to import and transform the data if there is a device failure due to error codes. Device failure is somewhat common on long deployments due to battery failure or water ingress and requires manually editing the error messages out of data files for them to process correctly. This function will contain one function for each file type that will not fail if there was an issue with the device but will instead read in all the data prior to the failure and print a warning containing information about the failure such as when it occurred, any available information on why it failed and the length of time that was recorded from initial deployment.

The Probablistic Algorithm

The core function (working title: 'prob_GLS_algorithm') is an extension of probGLS::prob_algorithm (Merkel et al. 2016) which uses a probabilistic algorithm to estimate the flight paths of seabirds. This package is primarily used by researchers in the northern hemisphere, particularly the North Atlantic/Europe, and as such is geared towards use in birds native to this area. The proposed function seeks to add functionality and remedy issues that occur when applying the algorithm to A) Areas outside of the usual zone of use and B) To species that cross the equator and thus have a significantly larger possible range of locations (bounding box) than other species.

Extension of Land Mask Functionality

The main use of `probGLS::prob_algorithm` is to track birds in relatively open habitats such as seabirds. This is due to the algorithm being sensitive to excessive shading of the light logger prematurely resembling sunset or alternatively, artificial lights leading to the miscalculation of sunrise. As the species are expected not to spend an appreciable amount of time far inland, particles are only generated over open water. In order to achieve this the function reads in an external file with location data for landmasses and reduces the chance of a particle occurring there to zero. One way to think of this is that the bounding box restricts location estimates to all the geographic points within that box and the land mask further restricts the possible locations to everything not denoted as land. A set of additional hardcoded land masks are implemented to restrict possibilities to areas outside of the Mediterranean, Baltic, Black and Caspian seas. This can be applied based on where a given species is known to travel and avoids skewing caused by implausible particle locations. These land masks are useful for birds native to Europe but having a larger set of land masks (e.g. the Atlantic ocean) or the ability to define one or more custom land masks is one of the goals of this package. To avoid adding an excessive number of arguments to the main function, a separate function will be used to define land masks and visualise the areas that they cover. The visualisation will help with troubleshooting incorrectly specified land masks before running the computationally intensive algorithm.

The above image is an example of prob_algorithm output for a group of Sooty Shearwaters. Note that due a combination of a large bounding box to cover flight patterns around the southern end of South America and the lack of custom land masks we see location estimates along the eastern seaboard of North America which based on what we know of the species is implausible. This illustrates the necessity of including additional or custom land masks to remedy this issue.

A Solution to Inter-Hemisphere Flight Paths

Given that a user has information about when migration occurs and therefore an idea of when a particular species will cross the equator this function aims to allow the user to define four arguments to specify the boundaries of two time windows:

- The earliest date of likely migration from the southern to northern hemisphere
- The latest date of likely migration from the southern to northern hemisphere
- The earliest date of likely migration from the northern to southern hemisphere
- The latest date of likely migration from the northern to southern hemisphere

where depending on how late in each window a time step is, a cloud is generated based on a bounding box that contains an increasingly small latitude range from the originating hemisphere. From the end of the first window to the beginning of the second, the bounding box will be restricted to the hemisphere that the bird is expected to be in. At the start of the second window the bounding box will return to its original specifications and the process will repeat. This method would allow each step to contain the same number of point estimates from which to calculate the median and should produce more reliable estimates.

Output

The output will be a object of class 'tracks' which will contain three elements:

- A list of the parameters used in the function
- A SpatialPointsDataFrame containing all the particle locations at each timestep
- A SpatialPointsDataFrame containing the most probable track (the median location of all points at each timestep)

By generating SpatialPointsDataFrame objects we can use plotting methods available to "sp" class objects but is easily readable and converted into a traditional dataframe if necessary.

Methods

Summarising

A generic summary call will generate the following information:

- A detailed list of the parameters used
- Summary statistics including:
  - Total duration of the data recorded.
  - Median, mean and standard deviation of distance traveled for both type 1 (Midday to midnight) and type 2 (Midnight to midday) instances. This could further be broken down by calendar month to better describe changes throughout the year.
  - Total distance estimated to be traveled over the deployment.

Visualisation

The object will have two visualisation methods, the first will show the location of all particles generated with a line overlapping which shows the median particle location at each step. The limits of the plot will be the bounds of the bounding box by default but can be modified to allow for a different perspective.

The second method will generate a series of plots, one for each step with appropriate labeling to denote which time step it is associated with. The user will have the ability to navigate through the plots and see all of the particles for that time step as well as the linked median points of all time steps. This will allow for a more detailed view of where each particle was generated and may be useful for troubleshooting issues such as the median track being significantly outside of the expected area at a given time. This view could inform the need for a new land mask to avoid significantly skewing the estimated location. The bounds of this plotting method will be the same as the original bounding box to allow for consistency and easy comparison between different plots. If a shifting bounding box was used because we expect an individual to cross the equator then the reduced bounding box will be drawn on each corresponding plot in the second visualisation method.

Data Requirements

Publicly available light logger data tends to be limited to species that have relatively short ranges and remain in a single hemisphere. Example data for the cross-hemisphere functionality will need to be sourced or synthetic data generated to showcase the functionality. If data can be sourced, permission to include it in the package will be required. If synthetic data is included then it must be explicitly mentioned in the documentation.

The package will also require the user to download the land mask data file and files relating to the ice concentrations and sea surface temperature for all the years that their deployment covers. The download location and required placement of this data will be noted in the documentation.

Testing

For the functions that read in and format the raw data from the devices there will be checks for the file type e.g. .trn for Biotrack/BAS and that the main section of the file has the correct number of comma delimited columns. Tests will be carried out to confirm that the columns are of the correct type. For a .trn file the first column is a datetime, followed by a string either 'Sunrise' or 'Sunset' and a numeric containing a status code.

Tests on the main 'prob_GLS_algorithm' will include:

- Checking that argument types are the correct type or coercible to the correct type.
- The order of the bounding box specifications is correct.
- The bounding box contains at least one non-land mask location.
- The output object is of the correct class.
- The child objects contained within the output object are both of the correct class and there is the correct number of them.
- The number of rows and columns in the child objects is consistent with the arguments provided to the algorithm.

Using a set of test data examples, performance tests will be carried out and the output of the tests will be checked against the output of `probGLS::prob_algorithm` using the same set of arguments and same random seed. If the new functionality isn't taken advantage of (i.e. there is no crossing of the hemispheres and no novel land masks are used) then we can expect the same particles to be generated at each timestep. The function that defines the land masks to use in the algorithm won't have a performance test but offers users a way to visually inspect if the result is satisfactory.

Ethical Considerations

The goal of this package is to make this type of tracking more accessible and user-friendly to researchers outside of Europe/the North Atlantic. This has the potential to encourage further use of this style of tracking device. There is a chance that the attachment of light loggers may impact the survivability of affected endangered birds but this risk was found to be lower than when using GPS trackers (Severson et al. 2019) and the data provided gives a greater understanding of movement patterns that could help to guide conservation efforts

Ekstrom, Philip A et al. 2004. "An Advance in Geolocation by Light."

Hill, Roger D, and Melinda J Braun. 2001. "Geolocation by Light Level." In *Electronic Tagging and Tracking in Marine Fisheries*, 315–30. Springer.

Merkel, Benjamin, Richard A Phillips, Sébastien Descamps, Nigel G Yoccoz, Børge Moe, and Hallvard Strøm. 2016. "A Probabilistic Algorithm to Process Geolocation Data." *Movement Ecology* 4 (1): 1–11.

Severson, John P, Peter S Coates, Brian G Prochazka, Mark A Ricca, Michael L Casazza, and David J Delehanty. 2019. "Global Positioning System Tracking Devices Can Decrease Greater Sage-Grouse Survival." *The Condor* 121 (3): duz032.

Wilson, Rory P. 1992. "Estimation of Location: Global Coverage Using Light Intensity." *Wildlife Telemetry: Remote Monitoring and Tracking of Animals.*