APPLICATION

FLightR: an R package for reconstructing animal paths from solar geolocation loggers

Eldar Rakhimberdiev[*,1,2] (iD)**, Anatoly Saveliev**[3]**, Theunis Piersma**[1,4] **and Julia Karagicheva**[*,1]

[1]*NIOZ Royal Netherlands Institute for Sea Research, Department of Coastal Systems and Utrecht University, P.O.Box 59, 1790 AB Den Burg, Texel, The Netherlands;* [2]*Department of Vertebrate Zoology, Biological Faculty, Lomonosov Moscow State University, 119991 Moscow, Russia;* [3]*Institute of Ecology and Environmental Science, Kazan Federal University, Kazan, Russia; and* [4]*Conservation Ecology Group, Groningen Institute for Evolutionary Life Sciences (GELIFES), University of Groningen, P.O. Box 11103, 9700 CC Groningen, The Netherlands*

Summary

1. Solar geolocators are relatively cheap and simple tools which are widely used to study the migration of animals, especially birds. The methods to estimate the geographic positions from the light-intensity patterns collected by these loggers, however, are still under development.

2. The accurate reconstruction of the annual schedules and movement patterns of individual animals requires analytical methods which provide estimates of daily locations, distances between the locations and the directions of movement, with measures of their uncertainty.

3. The new R package FLightR meets all these requirements. It enables refined and statistically validated estimations of movement patterns of birds. Here, we present main features of this advanced package.

Key-words: bird migration, geolocators, movement ecology, solar geolocation, template fitting, tracking

Introduction

Modern migration ecology builds on movement data of animals equipped with tracking devices (Nathan *et al.* 2008). Among the tracking devices, solar geolocators, i.e. archival tags that record light-level data over time, are now the smallest, cheapest and most widely used for migratory birds (Bridge *et al.* 2011, 2013). After their introduction to bird tracking about a decade ago (Croxall *et al.* 2005; Shaffer *et al.* 2006), data collected by these devices continue to dominate the field, with several tens of papers published in 2016. The greatest challenge in the use of solar geolocators is the stage of data processing: how are the sequences of light-level records that they store best converted into geographic positions (Lisovski *et al.* 2012b; Rakhimberdiev 2016)?

The reconstruction of the annual schedules and movement patterns is particularly challenging in animals with complex movements. Among these, migratory birds travel across a range of latitudes, with varying speeds and at different times of day (Alerstam 1990). The animals will use a variety of habitats, which may include sheltered ones, such as tropical forests. Here, recorded day lengths will be shorter than true day lengths, which will bias the estimation of geographic positions. Many seasonal movements along south-north axes occur during the equinoxes (Alerstam 1990), a time when day lengths are

equal all over the world. So just when it matters, the day lengths are uninformative. All these factors add to the imprecision of the results (Lisovski *et al.* 2012b).

An additional layer of complexity is generated by the need to not only know the geography of routes but also the timing of the movements, i.e. the schedules of individual animals (Marra *et al.* 2005; Conklin *et al.* 2010). We need to be able to distinguish between the states of being stationary and moving. Finally, to be confident about what can, and what cannot be concluded, we also should know the uncertainty of the estimates of place and time. The existing analytical tools for processing solar geolocation data, i.e. Trackit (Nielsen & Sibert 2007), TripEstimation (Sumner, Wotherspoon & Hindell 2009; Sumner & Wotherspoon 2016; currently replaced with SGAT), GeoLight (Lisovski, Hahn & Hodgson 2012a) and BAStag (Wotherspoon, Sumner & Lisovski 2016a) take care of some of the analytic demands, but never of all of them (see Rakhimberdiev *et al.* 2015 for review).

Here, we introduce FLightR, a new R (R Core Team 2016) package for analysis of light-level data obtained from solar geolocators. Compared with analogous packages, FLightR provides refined results in terms of accuracy and precision of calculated positions (Rakhimberdiev *et al.* 2016). The space and time estimates are improved by: (i) a hidden Markov chain model, computed via particle filter algorithm (Kitagawa 1996), and (ii) a template fit observation model allowing the use of all available measurements around twilights (Ekstrom 2004). The hidden Markov model provides better position estimates as it

*Correspondence authors. E-mails: eldar.rakhimberdiev@nioz.nl and julia.karagicheva@nioz.nl

utilizes not only positional information from the current twilight but also uses information from the previous and next twilights propagated through a movement model (Patterson *et al.* 2008). The template fit observation model relies on the principles of change in light irradiance with sun angle ('template') and is robust to varying degrees of shading (Ekstrom 2007).

FLightR works with light-level data collected by the various kinds of solar geolocators (Intigeo tags by Migrate Technology Ltd, GDL tags by Swiss Ornithological Institute), and particularly by the narrow light band tags (mk tags produced British Antarctic Survey, Biotrack Ltd and Migrate Technology Ltd). To avoid states of residence over water or over land surfaces in land- and seabirds, respectively, spatial and spatio-behavioural masks can be used. For every twilight, FLightR provides estimated positions of a tracked animal, the probability of its actually moving between twilights (over day and over night) and the distances and directions of movement. For the estimated positions, distances and directions, FLightR evaluates uncertainty. Rakhimberdiev *et al.* (2015) provide details on the package and compare its performance with other existing packages. FLightR is a cross-platform package, its core functions are implemented in parallel framework. FLightR is available on CRAN (https://CRAN.R-project.org/package = FLightR), the development version and examples of the use of this package are on GitHub at www.github.com/eldarrak/FLightR#readme.

Workflow

The workflow consists of the following steps: (i) annotation of twilights (done outside the FLightR package) and data import into the FLightR, (ii) calibration of the data, (iii) assignment of spatial extents, (iv) the setting up of priors and the compiling of the final data object, (v) running the particle filter analysis and (vi) visualization of the results. In Appendix S1, Supporting Information (also available on GitHub https://git.io/viTDY), we demonstrate the routine using a track of a black-tailed godwit (*Limosa limosa limosa*) migrating from its breeding site in the Netherlands to its wintering area in Spain. The raw dataset used is available from Rakhimberdiev *et al.* (2016).

DATA PREPARATION AND IMPORT

The analysis of geolocation data begins with the definition of the twilight events, i.e. the sunrises and the sunsets. As this procedure is implemented in several software packages, we did not incorporate it into FLightR. Instead, we recommend the use of the appropriate functions in other R packages, e.g. `preprocessLight` functions in BAStag (Wotherspoon, Sumner & Lisovski 2016a) and TwGeos (Wotherspoon, Sumner & Lisovski 2016b), `twilightCalc` in GeoLight (Lisovski, Hahn & Hodgson 2012a), or online service http://tags.animalmigration.org.

The FLightR uses the 'TAGS' format to read the annotated twilight data as input. This format represents an agreed general annotation of twilight data by the National Centre for Ecological Analysis and Synthesis working group 'Establishing an open-source animal-tracking analysis platform for archival geolocators' (https://www.nceas.ucsb.edu/featured/bridge). TAGS online service (http://tags.animalmigration.org) saves the data in the TAGS format. If one uses the R packages GeoLight, BAStag or TwGeos, the annotated twilights need to be exported to TAGS, for which the following functions in FLightR can be used: `GeoLight2TAGS`, `BAStag2-TAGS` or `TwGeos2TAGS`. The function `get.tags.data` reads comma separated file in TAGS format, detects tag type, checks whether the light data are log-transformed, transforms them back from the log scale if needed, and returns it as a FLightR data object.

CALIBRATION

Geolocator units measure light levels with varying precision, thus for each device calibration the relationship between the measured and the expected light levels needs to be established. This relationship can be estimated, using data, recorded during a so-called calibration period, i.e. when a tag stays in a known position, usually before the animal has left the area, where it is tagged or to where it returned the next year. The use of both periods (i.e. at the beginning and the end) helps to account for any ageing of the tag. For example, when an opacification of the transparent shell of a light sensor results in a decreased sensitivity of the device. The desired calibration period should at least be a few days. As a tagged animal will not always stay in a known location for a long enough period, we recommend 'rooftop calibration' (with a geolocator left outside in an open location) prior to, and preferably also after, tagging.

FLightR uses a 'template fit' for calibration (Ekstrom 2004). For each tag it finds the linear (on a log-log scale) relationship between the observed light levels (measured during the calibration period) and the theoretical light levels, estimated from current sun angle with the deterministic equation developed by Ekstrom (2004). During calibration at the known location for each twilight *i* with measurements *j*, FLightR estimates the slope Z_i of the log of the observed light level OLM_{ij} to the log of the expected light level ELM_{ij}:

$$\log(OLM_{ij}) = \alpha_i + Z_i \times \log(ELM_{ij}) + E_{ij}; \\ E_{ij} \in Norm(\mu_i, \sigma_i)$$

eqn 1

From the distribution of Z_i for all calibration twilights for given device, we then estimate the calibration parameters of a log-normal distribution:

$$Z_i \in logNorm(Z_{calib}, \sigma_{zcalib})$$

eqn 2

To evaluate likelihood of tag presence at a location *k*, using eqn (1) FLightR estimates local slope $\hat{z}_{ik}$ and its uncertainty $\sigma_{\hat{z}_{ik}}$. A joint probability of $\hat{z}_{ik}$, $\sigma_{\hat{z}_{ik}}$ and calibration parameters Z_{calib} and σ_{zcalib} used as the likelihood of occurrence of the animal at the grid point *k* during the twilight *i* (see Rakhimberdiev *et al.* 2015 for details).

The calibration parameters are based on the data recorded during calibration periods. The user needs to create a data

frame where the geographic coordinates of the calibration location, and the start and end dates of the calibration period, are specified. A calibration object is compiled with the function `make.calibration` from the data frame containing calibration periods and locations and the data object, obtained with the `get.tags.data`. The calibration object contains all necessary calibration parameters for the tag.

If the start and end dates are not known precisely, they can be visually determined. For this, the calibration slopes for sunsets and sunrises are obtained and plotted for every day of the tracking period based on the assumption that the tag remained at the same (calibration) location. Because calibration slopes reflect the adequacy of the light level measured by the device, as long as the tagged animal stayed put at the calibration location, they will vary little until the animal moves off. The function `plot_slopes_by_location` aims to help with the search for appropriate calibration periods.

In some cases, e.g. when animal is tagged during a 24-h polar day or when it migrates immediately after tagging, traditional calibration at a known site is not possible. With the function `find.stationary.location`, FLIGHTR determines the unknown locations where the animal has been resident. Here, the approach of FLIGHTR is similar to the Hill-Ekstrom calibration (Lisovski *et al.* 2012b) implemented in GeoLight (Lisovski, Hahn & Hodgson 2012a); the function searches for location at which the spread of errors in the slopes of twilights during resident period is minimal. This approach critically depends on a tagged animal being resident for at least few weeks.

ASSIGN SPATIAL EXTENT

FLIGHTR estimates the positions of an animal at twilights on a regular spatial grid with distance between nodes of approximately 50 km. The function `make.grid` sets up a spatial grid delimited by user-defined boundaries (the parameters `left`, `bottom`, `right` and `top`). Limiting the spatial extent of the grid will speed-up the calculations as it decreases the number of grid nodes that need to be processed. When the tagged animal simply cannot occur or stay between two subsequent twilights over particular areas, e.g. open water for a landbird, additional parameters `distance.from.land.allowed.to.use` and `distance.from.land.allowed.to.stay` can be specified. Each of these parameters requires a vector of two numbers: the minimal and the maximal distances (in km) from shoreline, at which the animal is allowed to occur/stay. For example, `distance.from.land.allowed.to.use = c(-100, 500)` means that the bird is allowed to fly in a range between 100 km inland, and up to 500 km from the shoreline over the sea.

PREPARE THE MODEL FOR RUN

The function `make.prerun.object` compiles a complex object to be used in the main run. This object incorporates all the four objects created at earlier steps: (i) the light data with the detected twilight events (Proc.data), (ii) the spatial parameters (Grid), (iii) the geographic coordinates of the location where the tracking has started (start) and (iv) the calibration parameters (Calibration). There are five parameters in the function (Bayesian priors for the movement model) that can potentially be manipulated: (i) the probability of migration (Decision, defaults to 0·1 for any transition), (ii) the direction of movement (Direction), (iii) the concentration parameter of von Mises distribution (Kappa, defaults to uniform distribution), (iv) the mean and (v) the SD of the distance that can be covered between subsequent twilights (M.mean, defaults to 300 km, and M.sd, 500 km). In case the relevant biological knowledge exists, prior values can be changed for all transitions, or for specified time periods.

PARTICLE FILTER RUN

The function `run.particle.filter` estimates the coordinates with credible intervals and probabilities of migration for each twilight event (runs ~ 45 min for godwit example on a standard workstation). For each twilight it also generates a probability of occurrence on the spatial grid, and a probability of transition between nodes. The output contains a table of positions at each twilight ($Results$Quantiles) and their statistics (mean, median values and credible intervals), a table of parameters of the movement model ($Results $Movement.results) and posterior distribution at every twilight and at every transition between twilights. At the end of the run, the function plots a simplified preliminary map with a resulting track, i.e. mean positions at each twilight, connected with a line.
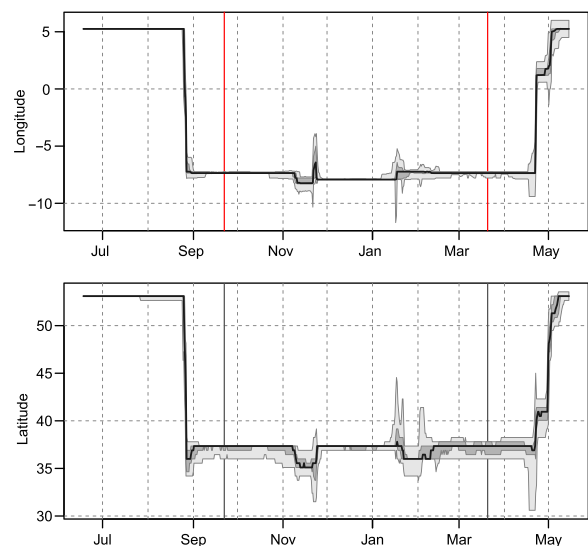

Fig. 1. Estimated longitudes (upper panel) and latitudes (lower panel) of animal positions, plotted with the function `plot_lon_lat (Result)`. The black lines represent the latitudes or longitudes with the median probability, the shaded areas represent latitudes or longitudes, in which probability of occurrence falls between first and third quartiles (dark-grey shade) and within the 95% credible interval (light-grey shade). The equinoxes are marked with red vertical lines.
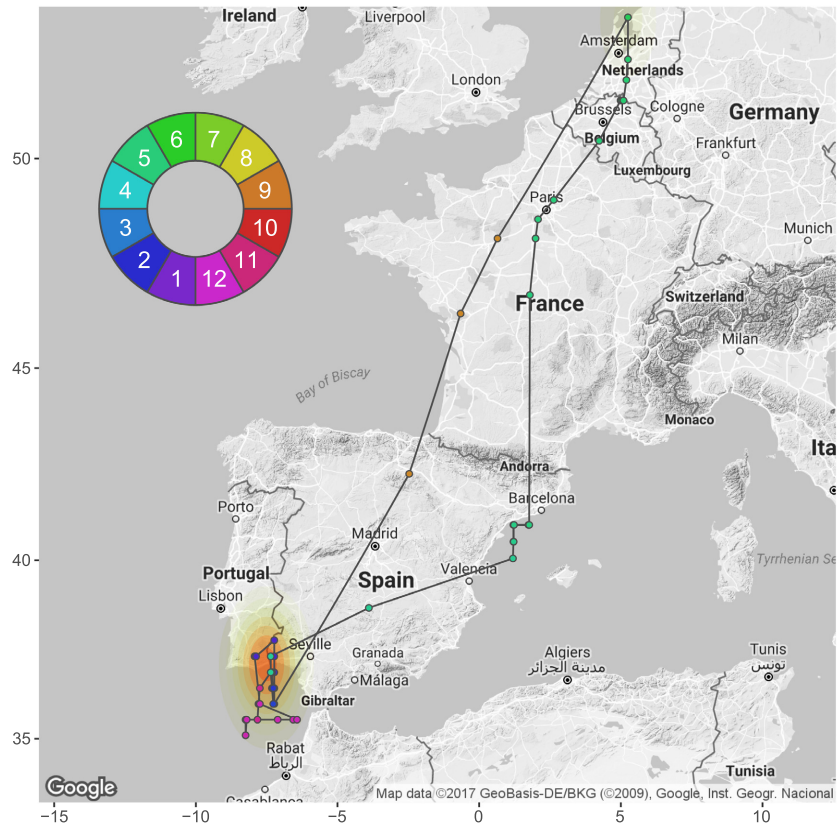
Fig. 2. A map of bird's movements plotted with the function `map.FLightR.ggmap (Result)`. The medians of estimated positions (by latitude and longitude) for each twilight are represented as coloured dots connected with lines. The colour of each dot corresponds to the colour of the months (on the colour-wheel), during which the twilight occurred. Shades of orange represent the utilization distribution. In the example, the bird spent most of the time in the wintering area.
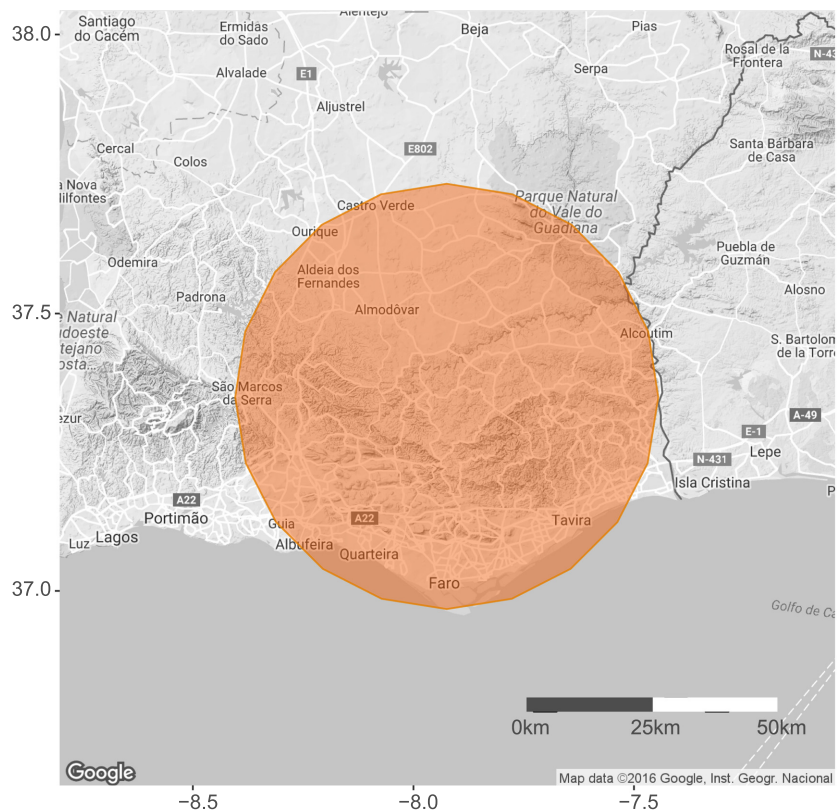

Fig. 3. A map with utilization distribution for the tracked bird, plotted with the function `plot_util_distr(Result, dates=data. frame(as.POSIXct('2013-12-01'),as. POSIXct('2014-01-31')), add.scale. bar=TRUE, percentiles=0·5)`. In this example, the defined area, indicated with orange shade, encompasses the grid point which accumulated more than 50% of probability of occurrence in December and January.

ESTIMATION OF ARRIVAL AND DEPARTURE TIMES

The function `find.times.distribution` derives the time at which an animal arrived or departed from an area of interest together with the measure of its uncertainty. A simple summary of the periods of residence and movement can be generated by the function `stationary.migration.summary`.

VISUALIZATION OF THE RESULTS

The results, i.e. the estimated probability of an animal occurring at a particular latitude and longitude on each day, can be plotted with the `plot_lon_lat` (Fig. 1). The function `map.FLightR.ggmap` provides a map with the most probable positions (combinations of the most probable latitude and longitude) for each day (Fig. 2). This function automatically detects the optimum spatial scale for the maps.

The function `plot_util_distr` is intended to plot utilization distributions, i.e. the probabilities of occurrence of an animal in space and time. It can also be restricted to a specific twilight or a group of twilights to visualize the most likely locations during certain periods such as the winter quarters or the migratory flights (Fig. 3). To do this, the initial and final dates of a period of interest are specified, e.g. the arrival and departure dates obtained during the procedure described in section 'estimation of arrival and departure times'.

Known limitations

FLIGHTR is well suited for most of the data obtained from solar geolocators, but there still are a few known limitations of its use. (i) FLIGHTR does not estimate latitudes well under high shading, e.g. for birds that are always under dense canopy. (ii) Narrow band mk tags have a limited amount of information when set for >5 min fixing interval. They still work well for low shading tracks, but fail earlier than other tags when shading increases. (iii) Imposed spatial masks restrict the effectiveness of particle filter and sometimes brings it to local minima, so we advise to always carefully compare results obtained with and without masks.

Conclusions

FLIGHTR is an advanced tool for the analyses of geolocation data. Compared with the existing solar geolocation packages it: (i) provides positions and migratory schedules of animals with a higher precision including estimates of their uncertainty, (ii) efficiently works with narrow light band tags, (iii) allows for systematic changes in the habitats used throughout an individual annual cycle, minimizing the bias from various shading regimes, (iv) estimates behavioural state parameters that describe a bird as either sedentary or on the move and (v) allows the modelling of movements of an animal over unsuitable habitats with restrictions on the animal staying there. These features enable the ecological interpretation of bird-tracking data.

Authors' contributions

E.R. and A.S. conceived methodology, E.R. implemented the methodology as an R package. J.K. and E.R. wrote package documentation. J.K. led the writing of the manuscript. T.P. collected the example data and contributed to the writing. All authors tested the software, contributed critically to the drafts and gave final approval for publication.

Acknowledgements

Data accessibility

FLIGHTR is available on CRAN (https://CRAN.R-project.org/package = FLightR) and GitHub (https://github.com/eldarrak/FLightR#readme).

References

Alerstam, T. (1990) *Bird Migration*. Cambridge University Press, Cambridge, UK.

Bridge, E.S., Thorup, K., Bowlin, M.S. *et al.* (2011) Technology on the move: recent and forthcoming innovations for tracking migratory birds. *BioScience*, **61**, 689–698.

Bridge, E.S., Kelly, J.F., Contina, A., Gabrielson, R.M., MacCurdy, R.B. & Winkler, D.W. (2013) Advances in tracking small migratory birds: a technical review of light-level geolocation: light-level geolocation dataloggers. *Journal of Field Ornithology*, **84**, 121–137.

Conklin, J.R., Battley, P.F., Potter, M.A. & Fox, J.W. (2010) Breeding latitude drives individual schedules in a trans-hemispheric migrant bird. *Nature Communications*, **1**, 1–6.

Croxall, J.P., Silk, J.R.D., Phillips, R.A., Afanasyev, V. & Briggs, D.R. (2005) Global circumnavigations: tracking year-round ranges of nonbreeding albatrosses. *Science*, **307**, 249–250.

Ekstrom, P. (2004) An advance in geolocation by light. *Memoirs of the National Institute of Polar Research, Special Issue*, **58**, 210–226.

Ekstrom, P. (2007) Error measures for template-fit geolocation based on light. *Deep Sea Research Part II: Topical Studies in Oceanography*, **54**, 392–403.

Kitagawa, G. (1996) Monte Carlo filter and smoother for non-Gaussian nonlinear state space models. *Journal of Computational and Graphical Statistics*, **5**, 1–25.

Lisovski, S., Hahn, S. & Hodgson, D. (2012a) GeoLight – processing and analysing light-based geolocator data in R. *Methods in Ecology and Evolution*, **3**, 1055–1059.

Lisovski, S., Hewson, C.M., Klaassen, R.H.G., Korner-Nievergelt, F., Kristensen, M.W. & Hahn, S. (2012b) Geolocation by light: accuracy and precision affected by environmental factors. *Methods in Ecology and Evolution*, **3**, 603–612.

Marra, P.P., Francis, C.M., Mulvihill, R.S. & Moore, F.R. (2005) The influence of climate on the timing and rate of spring bird migration. *Oecologia*, **142**, 307–315.

Nathan, R., Getz, W.M., Revilla, E., Holyoak, M., Kadmon, R., Saltz, D. & Smouse, P.E. (2008) A movement ecology paradigm for unifying organismal movement research. *Proceedings of the National Academy of Sciences USA*, **105**, 19052–19059.

Nielsen, A. & Sibert, J.R. (2007) State-space model for light-based tracking of marine animals. *Canadian Journal of Fisheries and Aquatic Sciences*, **64**, 1055–1068.

Patterson, T., Thomas, L., Wilcox, C., Ovaskainen, O. & Matthiopoulos, J. (2008) State–space models of individual animal movement. *Trends in Ecology & Evolution*, **23**, 87–94.

R Core Team (2016) *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

Rakhimberdiev, E. (2016) Ornithology by light levels today: dealing with a developing teenager. *Wader Study*, **123**, 1–3.

Rakhimberdiev, E., Winkler, D.W., Bridge, E., Seavy, N.E., Sheldon, D., Piersma, T. & Saveliev, A. (2015) A hidden Markov model for reconstructing

animal paths from solar geolocation loggers using templates for light intensity. *Movement Ecology*, **3**, 25.

Rakhimberdiev, E., Senner, N.R., Verhoeven, M.A., Winkler, D.W., Bouten, W. & Piersma, T. (2016) Comparing inferences of solar geolocation data against high-precision GPS data: annual movements of a double-tagged black-tailed godwit. *Journal of Avian Biology*, **47**, 589–596.

Shaffer, S.A., Tremblay, Y., Weimerskirch, H. *et al.* (2006) Migratory shearwaters integrate oceanic resources across the Pacific Ocean in an endless summer. *Proceedings of the National Academy of Sciences USA*, **103**, 12799–12802.

Sumner, M. & Wotherspoon, S. (2016). tripEstimation: metropolis sampler and supporting functions for estimating animal movement from archival tags and satellite fixes. version 0.0.44.

Sumner, M.D., Wotherspoon, S.J. & Hindell, M.A. (2009) Bayesian estimation of animal movement from archival and satellite tags. *PLoS ONE*, **4**, e7324.

Wotherspoon, S., Sumner, M. & Lisovski, S. (2016a). BAStag: basic data processing for light based geolocation archival tags. version 0.1.3.

Wotherspoon, S., Sumner, M. & Lisovski, S. (2016b). TwGeos: basic data processing for light-level geolocation archival tags. version 0.0.1.

Supporting Information

Details of electronic Supporting Information are provided below.

Appendix S1. FLightR analysis example of a black-tailed godwit geolocator.