APPLICATION

`GeoLight` – processing and analysing light-based geolocator data in `R`

Simeon Lisovski[1,2]* and Steffen Hahn[2]

[1]*Centre for Integrative Ecology, Deakin University, Geelong, Vic, 3220, Australia; and* [2]*Swiss Ornithological Institute, Seerose 1, Sempach, CH-6204, Switzerland*

Summary

1. Determining global position by light measurements ('geolocation') has revolutionised the methods used to track migratory birds throughout their annual cycle.

2. To date, there is no standard way of analysing geolocator data, making communication of analyses cumbersome and hampering the reproducibility of results.

3. We have, therefore, developed the `R` package `GeoLight`, which provides basic functions for all steps of determining global positioning and a new approach in analysing movement pattern.

4. Here, we briefly introduce and discuss the major functions of this package using example movement data of European hoopoe (*Upupa epops*).

Key-words: archival tag, bird migration, changepoint, geolocation, tracking

Introduction

The development of archival tags that can record geographical information through light intensity patterns ('geolocators') has greatly improved our knowledge of animal migration. Recent applications of light-weight geolocators (<2 g) have demonstrated their ability to investigate annual migration patterns of even small and clandestine bird species (Stutchbury *et al.* 2009; Bächler *et al.* 2010; Tøttrup *et al.* 2011; Bairlein *et al.* 2012; Schmaljohann *et al.* 2012) and have stimulated intended use of this methodology by many other researchers.

The accuracy of determining geolocation using light loggers relies on the accuracy of measuring times during sunrise and sunset. The most frequently used way to make these determinations is called the threshold method, whereby, sunrise and sunset times are identified as the time points when the light intensity passes a specific threshold. These time points, representing a given elevation angle of the sun, feed into standard astronomical equations that identify longitude and latitude (for details and background on geolocation see the study made by Hill & Braun 2001; Ekstrom 2004; Lisovski *et al.* 2012). As a consequence, any factor or process that affects ambient light levels may also influence the accuracy of determining a global position. Weather conditions and shading from vegetation are typically identified as the main factors compromising these measurements (Fudickar, Wikelski & Partecke 2011; Lisovski *et al.* 2012), but any physical or behavioural attribute that reduces light levels during critical periods may adversely affect

*Correspondence author. E-mail: slisovski@deakin.edu.au

data analysis. Thus, although the principle of geolocation is quite simple, the accurate analysis of the data is neither easy nor straightforward. This is further exacerbated when using ultra-light geolocators, which are constrained to record a restricted range of light levels because of their limited data storage and enabling the use of more precise positioning methods (e.g. template fit, see the study by Ekstrom 2004) and sophisticated analysing tools (Sumner, Wotherspoon & Hindell 2009; Pedersen *et al.* 2011). Therefore, we still lack a standardised procedure for analysing such data. Without a standardised analytical procedure, it is difficult to confidently compare results between studies.

We have, therefore, developed `GeoLight`, an `R` package for analysing light intensity data based on the threshold method. This analytical approach is applicable to all kinds of geolocator data and contains fundamental functions for every step of evaluating position: determination of sun events, discrimination of stationary and movement periods, calibration of these periods and, finally, calculation of positions.

Description

Geolocators record light intensity over time. These data can be loaded and processed with `GeoLight` using the following steps: (i) Determination of sunset and sunrise, (ii) Identification of stationary and movement periods, (iii) Calibration and (iv) Calculation of positions.

To demonstrate the package's major functions, we apply them to example data of a European hoopoe (*Upupa epops*) on autumn migration from Switzerland to Africa (Bächler *et al.*

2010). This data set is distributed as part of the `GeoLight` package (raw light intensity measurements: `hoopoe1`).

DETERMINATION OF SUNSET AND SUNRISE

The function `twilightCalc` determines sunrise and sunset as the times when the light intensity passes a particular threshold (Fig. 1). The function either requires a manually defined light intensity threshold (`LightThreshold`) or, in the default setting, uses a threshold of three light units above the baseline (the light intensity during the night). Ideally, the threshold is set within the twilight periods, where light intensities change most rapidly and shading has the least influence.

Sunset and sunrise times can be falsified by artificial light during the night or when, for example, a bird enters a nest box or a cave during the day. Therefore, the option `ask` in the `twilightCalc` function enables the user to confirm all automatically calculated sun events manually (and correct obviously erroneous assignments). The output of the `twilightCalc` function contains the basic values for almost all subsequent functions in `GeoLight`: a three-column data frame with `tFirst`, `tSecond` and `type`. The distributed data set `hoopoe2`, represents such a data frame and is derived by processing `hoopoe1` with the function `twilightCalc`.

```
> data(hoopoe2)
> hoopoe2[1:2,]
            tFirst            tSecond type
1 2008-07-15 03:13:00 2008-07-15 19:56:00    1
2 2008-07-15 19:56:00 2008-07-16 03:12:00    2
```

The first two columns represent two subsequent twilight events categorised by `type` (third column) as referring to a day (`type = 1`: `tFirst` for sunrise) or a night (`type = 2`: `tFirst` for sunset). The dependence of these values from the light intensity measurements is shown graphically in Fig. 1.

IDENTIFYING STATIONARY AND MOVEMENT PERIODS

`changeLight` is the function that distinguishes periods of residency and movement. To search for the time points during which the movement behaviour of an individual changed, we have implemented a changepoint model from the R package changepoint (Killick & Eckley 2011). Basically, such models assume data of an ordered sequence, $y_{1:n} = (y_1,...,y_n)$, which, in our case, are the ordered sunset or sunrise data. If a changepoint exists at a time $\tau \in \{1,...,n-1\}$, the statistical properties of two periods $\{y_1,...,y_\tau\}$ and $\{y_\tau,...,y_n\}$ are different. The procedure for identifying a single changepoint can be extended to look for multiple changepoints, m. The m changepoints will split the data into $m + 1$ segments, and each segment will be summarised by a set of statistical parameters. The implemented function `binseg.mean.cusum` uses a binary segmentation algorithm to efficiently search for changepoints by repeating the single changepoint method iteratively on different subsets of the time series (Scott & Knott 1974). To assess the magnitude of differences between the mean values of different segments, and therefore, the probability that the detected change is not caused by chance alone, a nonparametrical cumulative sum test (CUSUM) is used in the function `binseg.mean.cusum`. Rather than using final positions, the function `changeLight` relies on twilight events (e.g. defined sunrise and sunset times), which is advantageous for two reasons: (i) the procedure avoids inherent inaccuracies of determining positions (Hill & Braun 2001; Lisovski *et al.* 2012) and (ii) there are no data gaps around the equinoxes, thus, allowing the analysis of temporal migration pattern throughout the entire year.

To separate stationary periods, the function starts to search for time sequences between two changepoints with higher probabilities than user-defined thresholds for sunset and sunrise (Fig. 2a). Furthermore, the argument `days` can set the minimal duration before a stationary period is logged

```
> sites <- changeLight(hoopoe2$tFirst,
+ hoopoe2$tSecond,hoopoe2$type
+ rise.prob = 0.06,set.prob = 0.06, days = 5)
```

The calculated probabilities used to discriminate between the different stationary periods depend strongly on the degree of variance in sunrise and sunset times. Therefore, the changepoint probabilities might be higher in marine or open landscape environments than when species are in dense vegetation. The described variance may also differ for an individual between the two daily twilight periods (e.g. birds might be more active during one twilight period, thus, potentially result-
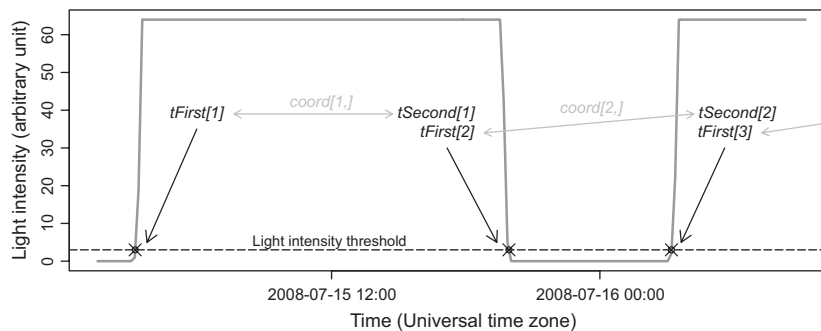

Fig. 1. First 36 h of light intensity measurements of the `hoopoe1` example (grey solid line) and the corresponding twilight events calculated by the function `twilightCalc`. Black labels and arrows are plotted accordingly to the output data frame from `twilightCalc`, whereas grey labels and arrows indicate the values used to derive geographical positions via the function `coord`.

ing in higher variance of either sunrise or sunset deviations). In this case, sunrise and sunset times can be analysed separately in `changeLight`. The user can then decide to use only one (e.g. `rise.prob=NA`) or both measurements to discriminate the stationary periods. Threshold values for sunrise and sunset probabilities may also differ (e.g. if the individual is more active during dawn than during dusk). There is an increased risk, however, in using very short values for the argument `days` (<4 days), as these may increase the likelihood of incorrectly assigning a stationary period. An example with two different

sunrise and sunset probability thresholds for the example data (`hoopoe2`) and using 5 days as the minimal duration period of residency is shown in Fig. 2.

CALIBRATION

Threshold-based positioning requires calibration, which can be achieved by using a reference sun elevation angle to fit the recorded day lengths to expected day lengths at a particular latitude at a particular time (for detailed information see Lisovski
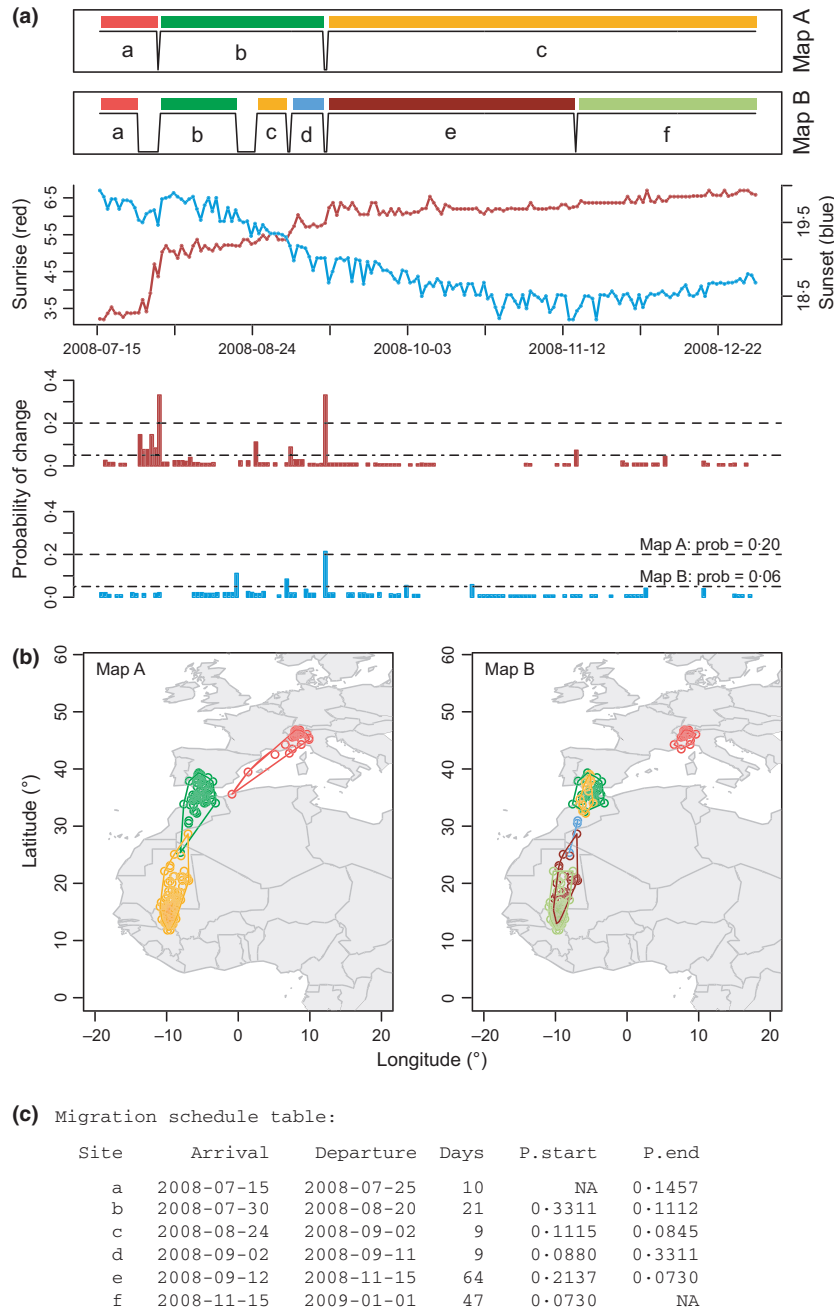
Fig. 2. (a) Slightly modified plot produced by `changeLight` for the analysis of stationary periods within the track. Here, two analyses with different defined thresholds (Map A: prob = 0·2; Map B: prob = 0·06) for the probability of change are compared. Probability thresholds are defined equally for sunrise and sunset within both analyses. (b) The two corresponding maps, made by `siteMap`. (c) The summary table of resulting migration pattern for Map B printed by the function `changeLight`.
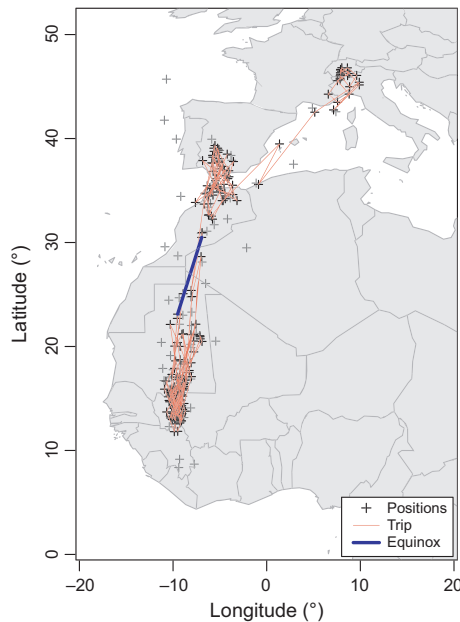
Fig. 3. Map produced by the function `tripMap` using the example data set `hoopoe2` with a `distanceFilter` (filtered positions are plotted in dark grey). The blue line combines the first and last point surrounding the equinox period where no latitudes could be derived by the function `coord` (in this case 36 twilight pairs).

et al. 2012 Supplementary Materials S2). Calibrations can, therefore, be used to partly account for shading. For example, in forest habitats, the recorded day length is expected to be shorter than the true day length, and thus, using a more positive sun elevation angle for these habitat measurements provide correct latitude positions (calibration does not affect longitudinal data).

The biggest challenge for calibration, however, is to find a suitable sun elevation angle for geologger data. One possibility is to record light data from a known site (usually the breeding site), and then, analyse these using `getElevation`. This function calculates the sun elevation angle, for which the median positions are closest to the known position and which reflects the shading conditions by habitat and weather during the stationary period. The breeding site of the European hoopoe used in this example is situated at 46·3°N and 7·1°W.

```
> getElevation(hoopoe2$tFirst[sites$site==1],
+hoopoe2$tSecond[sites$site==1],
+hoopoe2$type[sites$site==1],
+known.coord=c(7.1,46.3))
[1] -5.95
```

In this example data, the derived sun elevation angle is close to −6° ('Civil Twilight') – a sun elevation angle typical for species in open landscapes or marine environments with almost no shading from vegetation and during periods of clear weather conditions. It is important to recognise that the sun elevation angle can also be affected by the device architecture, in particular, the sensitivity of the light sensor.

The function `getElevation` may plot (ask=TRUE) sun elevation angles separately for sunrise and sunset. This analytical separation can be helpful if sunrise and sunset are affected

differently by shading, resulting in erroneous determination of day length and times of midnight/noon. In such case, a separate analysis might be helpful (for discussion about such effects see the study by Lisovski *et al.* 2012). By using the sun elevation angle identified from data at a known site (e.g. the breeding site, for the whole data set), we implicitly assume that shading patterns and thus, habitat use, major weather effects and behaviour were similar throughout the recording period. The extent to which this assumption is violated depends on the knowledge of the focal species and its ecology. In general, data derived in species using open landscape or marine habitats are less affected by shading factors and gives calibration a higher confidence.

The 'Hill–Ekstrom calibration' (Hill & Braun 2001; Ekstrom 2004; Lisovski *et al.* 2012) is also provided by the package `HillEkstromCalib` to refine selection of sun elevation angle. This function uses an iterative procedure that finds minima in the latitudinal variance over different sun elevation angles, separately for each stationary period. The underlying idea is that the error in latitude increases with an increasing mismatch between light level threshold and user-defined sun elevation angle. The function identifies the angle with the lowest latitudinal variance that fits best with the selected light intensity threshold and so results in the most accurate determination of position (see Fig. 1 in the study by Lisovski *et al.* 2012). As there might be several (local) minima, changing the starting angle (`start.angle`) in the iterations might yield different outcomes. The output provides the best angle for single or multiple stationary periods (`NA`'s will be produced if no latitudes can be calculated e.g. around the equinox, and if no minima in variance can be found between within −10° and 10°).

The principle idea behind the 'Hill–Ekstrom calibration' is simple – the user only needs to define stationary periods and search for the best sun elevation angle for each period. However, the method has strict requirements: (i) it can be applied to stationary periods only and (ii) pattern of shading must be stable throughout the whole stationary period. Although we can identify such periods rather easily and with a high degree of certainty by using the function `changeLight`, the second requirement is far more difficult to determine. Therefore, we suggest applying this method only for open landscape or marine species and/or for comparatively long stationary periods, thereby, minimising the adverse affects of irregular shading events.

POSITIONING

Geographical positions can be calculated with the function `coord`, which employs the standard astronomical equations (Montenbruck & Pfleger 2000) for converting two subsequent twilight events into corresponding latitude and longitude using the sun elevation angle determined in the calibration procedure (Fig. 1).

```
> positions<-coord(hoopoe2$tFirst,
+hoopoe2$tSecond,hoopoe2$type),
+degElevation=-5.95
```

```
> Note: Of 340 twilight pairs, the calculation of 62
latitudes failed (18%)
```

The notation in the function output refers to the inherent problem of geolocation by light; namely that position cannot be calculated at latitudes having almost equal day length (equinox periods). For such periods and for regions with almost no twilight events (i.e. polar regions), the function will produce NA values in coord[,2].

Furthermore, changes in the animal's behaviour during twilight periods (e.g. by habitat change) or extreme shading at particular dates can result in prediction of unrealistic positions. Some of these problems can be overcome using the function distanceFilter, which can eliminate some obvious outliers by using a maximal distance that an individual can cover in a given time

```
> filter <- distanceFilter(hoopoe2$tFirst,
+hoopoe2$tSecond, hoopoe2$type,
+degElevation=-5.95, distance=30)
> Note: 36 of 278 positions were filtered (22%)
```

In our example, the hoopoes are not expected to fly faster than 30 km h^{-1} (distance = 30) for the entire period between two subsequent twilight events (Bächler *et al.* 2010).

VISUALISING DATA

tripMap is a simple mapping function providing an overview of the resulting positions and their temporal order along with a bold line that indicates data gaps around the equinox (Fig. 3). The second mapping function siteMap gives an overview of the identified stationary periods (Fig. 2b). For a first impression, a convex hull shows the maximal distribution/area of each site. Nevertheless, further statistical techniques can be used (e.g. kernel density estimation) and need to be refined in order to describe the positions or the area in a more correct manner, for example, by accounting the seasonal and spatial variability in the accuracy of positioning.

Conclusion

The main objective of the GeoLight package is to provide an analytical tool that will benefit a wide range of geolocator users. All functions are described in detail in the R Package help documents, as well as all arguments and options for this specific application. In some cases, the context of using this method and its underlying assumptions are raised.

We consider GeoLight as a first step towards standardising geolocator analyses, which should facilitate better reproducibility and communication of analysis and results. We welcome feedback, comments and suggestions on GeoLight and appreciate all further developments that might advance the analysis of light-based geolocator data. We have created a discussion forum to facilitate communication and to encourage dialogue for further advancing these methods (http://simeonli sovski.wordpress.com/geolight/). The package is freely available on CRAN (cran.r-project.org).

Acknowledgements

References

Bächler, E., Hahn, S., Schaub, M., Arlettaz, R., Jenni, L., Fox, J.W., Afanasyev, V. & Liechti, F. (2010) Year-round tracking of small trans-Saharan migrants using light-level geolocators. *PLoS One*, **5**, e9566.

Bairlein, F., Norris, D.R., Nagel, R., Bulte, M., Voigt, C.C., Fox, J.W., Hussell, D.J. & Schmaljohann, H. (2012) Cross-hemisphere migration of a 25 g songbird. *Biology Letters*, **8**, 505–507, doi: 10.1098/rsbl.2011.1223.

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

Fudickar, A.M., Wikelski, M. & Partecke, J. (2011) Tracking migratory songbirds: accuracy of light-level loggers (geolocators) in forest habitats. *Methods in Ecology and Evolution*, **3**, 47–52.

Hill, C. & Braun, M.J. (2001) Geolocation by light level – the next step: latitude. *Electronic Tagging and Tracking in Marine Fisheries* (eds J.R. Sibert & J. Nielsen), pp. 315–330. Kluwer Academic Publishers, The Netherlands.

Killick, R. & Eckley, I.A. (2011) changepoint: An R package for changepoint analysis. R package version 0.5. http://CRAN.R-project.org/package=changepoint.

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

Montenbruck, O. & Pfleger, T. (2000) *Astronomy on the Personal Computer*. Springer, Berlin.

Pedersen, M.W., Patterson, T.A., Thygesen, U.H. & Madsen, H. (2011) Estimating animal behavior and residency from movement data. *Oikos*, **120**, 1281–1290.

Schmaljohann, H., Buchmann, M., Fox, J.W. & Bairlein, F. (2012) Tracking migration routes and the annual cycle of a trans-Sahara songbird migrant. *Behavioral Ecology and Sociobiology*, **66**, 915–922.

Scott, A.J. & Knott, M. (1974) A cluster analysis method for grouping means in the analysis of variance. *Biometrics*, **30**, 507–512.

Stutchbury, B.J.M., Scott, C.A., Done, T., Gow, E., Kramer, P.K., Tautin, J., Fox, J.W. & Afanasev, V. (2009) Tracking long-distance songbird migration by using geolocators. *Science*, **323**, 896.

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

Tøttrup, K.P., Klaassen, R.H.G., Strandberg, R., Thorup, K., Kristensen, M.W., Jorgensen, P.S., Fox, J., Afanasyev, V., Rahbek, C. & Alterstam, T. (2011) The annual cycle of a trans-equatorial Eurasian-African passerine migrant: different spatio-temporal strategies for autumn and spring migration. *Proceedings. Biological Sciences/The Royal Society*, **279**, 1008–1016.