RESEARCH ARTICLE

Twilight-free geolocation from noisy light data

Aidan D. Bindoff1 | Simon J. Wotherspoon2,3 | Christophe Guinet4 | Mark A. Hindell5,6

1Wicking Dementia Research and Education Centre, University of Tasmania, Hobart, Tasmania, Australia; 2Institute for Marine and Antarctic Studies, University of Tasmania, Hobart, Tasmania, Australia; 3Australian Antarctic Division, Kingston, Tasmania, Australia; 4Centre d’Etudes Biologiques de Chizé, Centre National de la Recherche Scientifique, Villiers en Bois, France; 5Institute for Marine and Antarctic Studies, Hobart, Tasmania, Australia and 6Antarctic Climate & Ecosystems Cooperative Research Centre, University of Tasmania, Hobart, Tasmania, Australia

Abstract

1. Solar geolocation is used to quantify the movements of animals tagged with sensors that record ambient light with respect to time. Global location sensor (GLS) tags are small, light, and present minimal drag or wing loading. They are affordable, and some can record data for several migratory cycles. These benefits mean they can be used in applications for which satellite tags are unsuitable. However, large errors in estimated locations can result if the sensor is obscured, especially around twilight, and sometimes the data obtained is unusable by existing methods of analysis due to this source of noise. This places limitations on the usefulness of solar geolocation in conservation and monitoring efforts.

2. All existing methods of analysis are dependent on twilights being identifiable or faithfully recorded. Instead, the method introduced here depends on the overall pattern of day and night to calculate the likelihoods for a Hidden Markov Model, where the hidden states are geographic locations. We call this a “twilight-free” method of light-based geolocation.

3. This method quickly estimates locations from otherwise unusable noisy light data. We use examples to show that the method produces tracks that are comparable in accuracy and precision to other geolocation methods. Furthermore, efficiency and replicability of estimated paths are improved because the user does not have to subjectively identify twilights. Other data sources, such as sea surface temperature and land or sea masks are easily incorporated, further improving location estimates and processing speed.

4. The twilight-free method offers new opportunities to researchers interested in the movements of animals that routinely have obscured sensors, or to analyse previously unusable noisy light data. It offers a fast, efficient, and replicable method for analysing tag data without the need for time-consuming pre- or post-processing. By increasing the yield of usable data from GLS tagging studies, researchers can more efficiently quantify where animals are going and when, and monitor changes in habitat. This is of fundamental importance to management and conservation efforts.

KEYWORDS

animal tracking, archival tag, bird migration, global location sensor, Hidden Markov Models
1 | INTRODUCTION

Solar geolocation is the process of inferring geographic locations from recorded patterns of day and night. In biology, it is used to reconstruct the large-scale movements of animals tagged with sensors that record ambient light, known as global location sensor (GLS) tags (Hill, 1994; Hill & Braun, 2001). Advances in technology have enabled the production of GLS tags that weigh as little as 0.5 g (Bridge et al., 2013) and are no bigger than a thumbnail. This enables tags to be deployed on a wide range of migratory species, even small songbirds (Bridge et al., 2013; Stutchbury et al., 2009). The small size, affordability, and long battery life of GLS tags make them suitable for applications that require more energy intensive satellite tags are not suitable for. One or more migratory cycles can be tracked in a single deployment, revealing important foraging habitats and migratory corridors, and without the attachment issues, expense, or weight penalty of satellite tags. However, unlike alternative techniques such as GPS, reconstructing movements from GLS tag data is an indirect method in which locations need to be estimated, using light intensity and time as geographical proxies, with the aid of computer software.

Current methods of solar geolocation can be classified into two broad types: threshold and curve/template-fit methods. Threshold methods estimate the time of twilight as the time at which the recorded light crosses a pre-determined threshold (Ekstrom, 2004; Hill, 1994). Locations are then derived from successive pairs of twilights. In essence, longitude is determined from the time of apparent noon and midnight, and latitude from apparent day length. Curve/template-fit methods (Ekstrom, 2004; Sumner, Woltherspoon, & Hindell, 2009) rely on the fact that the boundary between light and dark sweeps across the earth faster at the equator than at the poles. This has the consequence that, to an observer, the sun appears to rise and set faster at the equator. Curve/template-fit methods estimate longitude from the time of twilight and latitude from the rate of change in light near twilight.

A necessity of both methods is that ambient light at twilight is faithfully recorded, and it follows that solar geolocation methods are sensitive to noise from sensor obscuration or shading in the light data. The ability to estimate accurate locations from GLS data is severely hampered by obstruction of light to the sensor: e.g., sensor shading or obstruction caused by roosting, diving, returning to nests or burrows, feathers, vegetation, the landscape, or heavy cloud cover (Fudickar, Wikelski, & Partecke, 2012; Hill & Braun, 2001; Lisovski et al., 2012; Musyl et al., 2001; Welch & Eveson, 1999). Figure 1 shows a daily light record from two tags. The upper record is relatively noise free and it is a simple matter to identify times of twilight, sufficient for a good estimate of where the animal was on that day. The lower record shows an example where the sensor had been obscured for much of the day, and it is difficult to conclusively identify the true period of daylight.

Even relatively small errors due to noise can contribute to large effects on location estimates. Each minute of error in estimating a single twilight will result in an error of approximately 0.25 degrees in longitude (Hill, 1994). Errors in latitude can be considerably worse, particularly around the equinox (Hill & Braun, 2001), where due to the angle of the earth in its solar transit, day length becomes increasingly similar at all latitudes. As a consequence, small estimation errors from the light data can cause large errors in location estimates, particularly of latitude.

Outliers in light data can sometimes be detected by visual inspection, and either given a low confidence score, corrected, or data for the day excluded from analysis. This is time consuming and subjective, compromising replicability (Bridge et al., 2013). Objectivity and replicability may be improved by utilising automated procedures. Nonetheless, often twilights cannot be identified from very noisy light data, and the tracks produced using poorly estimated twilights are likely to be implausible or misleading. The expected benefits of affordable, large sample-size tagging programmes may go unrealised because the resulting data are compromised by noise.

In response to the need for a solar geolocation method that is robust to substantial noise from obscuration in the light data, we present a new method that relies on the entire light record rather than an explicit dependence on twilights. Because it is not necessary to identify twilights using this method, we call it a “twilight-free” method of solar geolocation. The method requires minimal user input, thus improving replicability and reducing labour. The method uses computationally efficient algorithms to arrive quickly at the best estimate of locations from the observed data without the necessity of post-processing. Being robust to noise, the yield of usable tracking data is improved, saving valuable time and effort. Furthermore, it is capable of incorporating other data such as sea surface temperature (SST) into the location estimation process, in a Bayesian framework. We illustrate the twilight-free method by reproducing animal tracks from heavily obscured light data. The first example uses a simulated light and SST record based on an existing short-tailed shearwater Ardenna tenuirostris (Temminck, 1835) track, and then we produce several tracks using data from GLS tags recovered from southern elephant seals Mirounga leonina (Linnaeus, 1758) that also carried GPS tags for validation. Parameter sensitivity is analysed in an informative manner for a typical user application.
2 | MATERIALS AND METHODS

The twilight-free method frames the geolocation problem as a Hidden Markov Model (HMM) (Rabiner, 1989) in which the hidden states are the daily geographic locations and the measured response is the observed pattern of light and dark over the day.

A Hidden Markov Model assumes that at a sequence of discrete times $t = 1, 2, \ldots T$ a system is in one of $C$ discrete states. Transitions between states are assumed to obey the Markov property, that is, the probability of the system being in state $s_{t+1}$ at time $t + 1$ depends on the state $s_t$ of the system at time $t$, but is otherwise independent of any previous state.

$$
Pr(s_{t+1} = j|s_t = i) = p_{ij}
$$

The state of the system is not observed directly, but at each time $t$ an observation $d_t$ is taken, the distribution of which depends on the hidden state $s_t$

$$
d_t|s_t \sim F(t)
$$

and optionally a set of unknown parameters $\theta$.

Given observations $d_1, d_2, \ldots, d_T$ and a prior distribution $p(s_1)$ for the initial state, inferences are drawn on the hidden states $s_1, s_2, \ldots s_T$. For a closely related example, if we do not know our location but we do know the time and the position of the sun in the sky, we can infer our location from these observations. The hidden states in this example are the locations, and the observations are the time and position of the sun in the sky.

To cast the geolocation problem as a HMM, the spatial domain is discretised to form a grid of cells and time is discretised into successive segments of 24 hr. It is assumed (for convenience) that the animal occupies a single cell in each 24 hr segment, and so the state $s_t$ represents the cell in which the animal is located in the 24 hr segment $t$, and the observed measurement $d_t$ is the entire light record for the 24 hr segment. The hidden states, $s_1, s_2, \ldots s_T$, are the unobserved locations estimated by the HMM.

The likelihood for each state, $s_t$, is determined by a comparison between the observed light record, $d_t$, and the expected pattern of day and night at each cell. Standard astronomical formulae (Meeus, 1991) are used to determine when each cell would be in daylight. Specifically, at each time that there is a sample of recorded light, solar zenith angles are calculated for the centre of each cell. When the solar zenith angle is above a calibrated threshold (typically 95–98°) we do not expect the cell to be illuminated. If any light is observed when the cell is expected to be in darkness the likelihood of the animal’s position being inside that cell is assumed to be zero. Where the sensor records darkness when the cell is expected to be illuminated, it is assumed that light to the sensor has been obscured. The total time $\tau$ that the sensor is assumed to have been obscured is calculated and the likelihood of the animal’s position being inside that cell is derived by assuming $\tau$ to follow some distribution, we use a Gamma distribution for flexibility,

$$
(\tau|s_t = i) \sim \text{Gamma}(\alpha, \alpha_2)
$$

The transition probabilities describe the probability of moving from one cell (or state, as here they are used interchangeably) to another on successive days

$$
p(s_{t+1} = j|s_t = i) = p_{ij}(t)
$$

These $p_{ij}(t)$ are then determined by assuming the distribution of average daily travel speeds are Gamma distributed where if $D_{ij}$ is the distance from cell $i$ to cell $j$ and $\Delta t = 24$ h is the number of hours in a segment

$$
\frac{D_{ij}}{\Delta t} \sim \text{Gamma}(\beta_1, \beta_2)
$$

with $\beta_1$ and $\beta_2$ chosen to reflect the distribution of average speed. There is no particular physical process that supports the choice of a Gamma distribution, other than non-negativity and inherent flexibility. In the fullness of time, other distributions might prove to offer even better approximations.

Hidden Markov Models are fitted by a recursive two pass forward-backward algorithm (Zucchini & MacDonald, 2009). A first pass propagates information forward through time, the probability of being in state $s_t$ at time $t$ is recursively computed by summing all pathways to that state from all possible states at the previous time step. Similarly, a second pass propagates information backwards in time, recursively summing over all pathways from that state to all possible states at the subsequent time step.

A classical HMM typically has comparatively few hidden states, whereas the geolocation problem has an enormous number of states, one for each cell of the grid. To make the forward-backward algorithm computationally feasible, for each time step our algorithm identifies cells with low relative likelihood and excludes them from the computation at that step. Essentially this treats the low probabilities as if they are zero, excluding the most unlikely locations from consideration. The centre co-ordinate of the cell with the maximum a posteriori estimate (Gauvain & Lee, 1994) is taken to be the inferred location estimate for each day.

In order to estimate which observations dichotomously represent “light” and which observations represent “dark,” and the range of solar zenith angles that represent “day” and “night,” it is important to calibrate tag data. Ideally the sensor would record light at a known location (or locations) for several days so that the tag light threshold and zenith angle that best represent twilight can be determined (Ekstrom, 2004; Hill, 1994).

The method is easily extended to incorporate other available information such as sea surface temperature (SST), land/sea masks, and sightings from fixed locations to further improve location estimates. Sea surface temperature observations from tag data can be compared to remotely sensed SST observations, giving improved estimates of latitude (Delong, 1992; Nielsen, Bigelow, Musyl, & Sibert, 2006; Teo et al., 2004). For animals that are constrained to land/sea, the prior probability of being on sea/land can be set to zero, eliminating those cells from consideration. Similarly, priors can be chosen to favour movement where it is most likely, but allow movement where it is possible but unlikely. Where there has been
a sighting at a fixed location, \( c \), at time \( t \), the likelihood for all other cells is zero. A further benefit of using all available information is in reducing computation time by reducing the number of cells under consideration.

An implementation of the method using the R language (R Core Team, 2016) is provided at www.github.com/ABindoff/geolocationHMM. Complete datasets for two examples below are provided, along with code to replicate one of the southern elephant seal tracks (tag 86372) and verify the accuracy and precision statistics reported in this paper against the reference track. Tools are provided to enable users to find optimal movement and shading parameters for their species of interest, should double-tagging data be available.

### 2.1 Example 1: simulated noisy light record of a short tailed shearwater migration

Migratory bird species display a range of movement-related behaviours over a large global domain, including foraging, breeding (egg incubation and chick-rearing), and migration. In order to demonstrate the utility of the twilight-free method in revealing these characteristic movement patterns and when they occur, a light record was simulated from positions taken from a short-tailed shearwater \( At. tenuirostris \) track, estimated previously using the method of Sumner et al. (2009).

This track covered the full migratory cycle of the short-tailed shearwater (STSH), from southern to northern high latitudes including two equinoxes and equatorial passages. Behaviours included incubation, chick rearing, and migration. Adding to the challenge, this species tends to migrate during equinoxes.

Remotely sensed SST data were appended to the light record at regular intervals. Noise (intermittent periods of darkness drawn at random from an exponential distribution) was added to a duplicate of this light and SST record. This treatment made it impossible to identify true twilights from the light record (Figure 2), and would certainly introduce errors in position estimates using threshold methods. A summary of the simulation data including sampling rate is given in Table 1. The “noisy” and “noiseless” light records were then processed and the resulting tracks estimated.

### 2.2 Example 2: double tagged southern elephant seals

The frequent, deep dives of the southern elephant seal \( Mirounga leonina \) present issues for estimating positions from a light record, because light to the sensor is attenuated at depth, but attachment issues prevent longer-term tracking of animals using satellite telemetry. We used light records from an adult female from Isles Kerguelen who carried both a GLS tag (Wildlife Computers TDR-Mk 9) and a GPS recorder (Wildlife Computers Fastloc-GPS). Depth and light were sampled every 2 s, and then were down-sampled to two minute intervals by taking the maximum observed light in each two minute interval (a more complete summary of the data is given in Table 1). The GLS tags were attached to pre-breeding adult female at Isles Kerguelen following the method of Dragon, Bar-Hen, Monestiez, and Guinet (2012). The mean dive frequency was approximately 2.7 dives per hour with a mean depth of 463 m and minimal surface time, meaning that light attenuation due to diving was very common. Although sea surface temperature data were available, these were excluded from the analysis to demonstrate performance of the model using light alone.

Errors in GPS positions were filtered using the “trip” R package (Sumner & Luque, 2016), which uses the speed filter described by McConnell, Chambers, and Fedak (1992). Threshold and solar zenith angles were calibrated, using a single day of light and GPS data from the beginning of each trip. These were approximate because the animal moved during the period used for calibration.

### 2.3 Parameter selection and sensitivity

There were substantial differences in the cell transition and shading likelihood parameters used in the models (Table 1). The underlying intuition is that the cell transition distribution should restrict animal movement to likely positions, and the shading likelihood distribution should reflect the likelihood that observations on the daylight side of twilight are shaded for \( k \) minutes. In effect, it is this shading that makes one cell a more or less likely candidate than another if the observed light could have occurred in either. Accordingly, we could safely assume that there was no shading in the “noiseless” short-tailed shearwater dataset, while we estimated that the “noisy” dataset was shaded for some length of time at \( \frac{1}{4} \) of all twilights. We assumed that the tag sensors were shaded at nearly every twilight for the southern elephant seals. Reasonable estimates of shading must suffice, as the true degree of shading is usually unknown.
Similarly, the true distribution of movement speeds is unknown and a reasonable estimate from prior knowledge is sufficient. The cell transition parameter restricts unreasonable movements, but it doesn’t dictate where or how far the animal moves. Prior knowledge from previous tracks, double-tagging a subset of animals with a satellite tag (where possible), or other information about the animal’s movement patterns can be used to inform cell transition estimates. For example, we know that STSH travel quickly during migration and so our cell transition parameters had to allow for this movement through a long-tailed distribution, whereas tighter restrictions were placed on SES movements.

For the SES datasets, 5 movement × 5 shading parameter combinations were tested, using data from a single animal (tag 86372). Parameters for the most accurate (minimising average great circle distance from the reference positions) were then used to compute tracks for the remaining animals tags 86373 & 78524). This analysis is detailed (including R code) at www.github.com/ABindoff/geolocationHMM.

### 3 | RESULTS

The method succeeded in producing tracks of comparable or better accuracy and precision to other methods, for example Phillips, Silk, Croxall, Afanasyev, and Briggs (2004), Shaffer et al. (2005), and Teo et al. (2004). Figures 3, 4, and 5 show some of the plotted estimated tracks overlaid with the reference tracks. The proximity of estimated positions to the reference positions are summarised in Table 2.

The short-tailed shearwater locations have been grouped into “resident,” “migratory,” and “resident and migratory” periods for both “noisy” and “noiseless” simulations. “Resident” periods refer to periods where the bird is engaged in incubation, chick rearing, or foraging in a well-identified area. “Migratory” periods refer to periods where the bird is moving between “resident” habitats (e.g., equatorial passages). “Resident and migratory” refers to the whole track inclusive of both periods. Accuracy and precision of location estimates during “resident” periods was higher than during “migratory” periods for both “noisy” and “noiseless” conditions. The mean great circle distance (gcd) from the GLS location estimate to the simulated true position on each day was 119.6 km (SD = 95.0 km) for the “noisy” condition and 90.0 km (SD = 75.1 km) for the “noiseless” condition. The simulated track had a fix at each twilight so the average longitude and latitude were taken for comparison.

Interestingly, the accuracy and precision of the method during migratory periods was better for the “noisy” simulation than the “noiseless” simulation, with a mean gcd of 554.0 km (SD = 706.6 km) and 798.3 km (SD = 1,339.6 km), respectively. The method failed to find a solution around the equinox on March 20th, so data for March 15th to March 27th were excluded from the analysis. Excluding a
similar period around the equinox on September 22nd improved results considerably, with a mean gcd of 425.0 km (SD = 458.9 km) and 416.5 km (SD = 449.6 km) for the “noisy” and “noiseless” conditions respectively. The magnitude of the errors are related to the distance travelled by the animal, and the true mean daily movement of the animal over the “migratory” periods was 687.6 km (SD = 680.7 km). Overall performance of the method in terms of accuracy and precision for “noisy” and “noiseless” conditions was comparable.

A similar result to the “foraging” periods of the STSH was obtained for all southern elephant seal tracks, with a mean gcd to the nearest filtered GPS fix on each day of 73.3 km (SD = 44.8 km)—116.0 km (SD = 57.5 km). The nearest GPS fix was taken as the reference for comparison in order to reduce some of the error attributable to movement alone (providing a measure of how well the method estimates positions that the animal actually visited). A full summary of the parameter sensitivity analysis is provided in supplementary materials at www.github.com/ABindoff/geolocationHMM, and shows that even quite large deviations from the optimal parameters produce acceptable tracks, sufficient to identify important foraging habitats and time spent within those habitats.

Density plots (Figure 6) indicate that the distribution of distances covered each day was well estimated by the method for both species. Importantly, foraging and nesting habitats (which the STSH return to often) are well identified, along with migratory routes, and times of departure and arrival.

4 | DISCUSSION

We have shown that a new twilight-free method for estimating animal movements from recorded light data provided credible tracks from otherwise potentially unusable noisy light data. Without the need for users to estimate time of twilights, the method is objective and the estimated tracks are replicable. The method is reasonably insensitive to departures from optimal shading and movement parameters used in the estimation of location likelihoods.
Although robust to noisy light data, the method has no in-built mechanism to deal with noisy SST data so care must be taken to ensure that instruments are calibrated and that observations are of SST and not the temperature of deeper water or body warmth. Nevertheless, the SES tracks demonstrate that SST data is not necessary to reproduce informative tracks.

Similarly, the method is not robust to light pollution from other light sources, such as ships, lighthouses and human activities. The likelihood is calculated assuming the sun is the only source of light the tag can record. If the tag encounters non-solar light sources at night, locations that are reasonable might be excluded. Where this light pollution occurs clearly within the night, removal is straightforward and justifiable; however, where light pollution occurs near twilights, removal becomes more subjective. Special consideration should be given for birds who can fly into and out of polar circles, thus observing periods of light several times per day. These observations are not light pollution, and are informative in this model.

Although this method does not require the user to identify twilights and is robust to noise in the data, good tracks cannot be reconstructed if the properties of the GLS tag sensors are not known. Regardless of the method, sound calibration procedures should be employed so that the apparent tag threshold and solar zenith angle can be determined with precision. If in doubt, the threshold can be set a little high (because the method is robust to noise), but informative observations may be lost.

Spatial discretisation necessarily introduces error. Spatially continuous methods, such as SGAT (Sumner et al., 2009) or Geolight (Lisovski & Hahn, 2012), avoid this problem but suffer greatly when errors are made estimating twilights. Even without the sensor being shaded, a sampling rate of 4 min will yield errors of up to 1° longitude at each twilight, which is more than the error expected as a consequence of using a 1° × 1° grid. In effect, the nett error from spatial discretisation is comparable to the error due to temporal discretisation during data collection.

Another consequence of discretisation is that any true movement within a single cell from day t−1 to t will be estimated (correctly, according to the model) as no movement (the animal is estimated to be at the centre of the cell on both days). Hence, any continuous distribution used to induce cell transition probabilities that has a probability mass of zero at a speed of zero will force movement out of the cell. Hence, we used a shape parameter of 1 (an exponential distribution in other contexts) because this was a simple way to restrict movement to a reasonable range without forcing unrealistic movements when the animal did not range far. Other distributions (e.g., truncated normal) may further improve location estimates.

Conversely, large movements out of the cell introduce other potential errors. With all methods of light-based geolocation, movements in an easterly direction result in shorter apparent day length, and movements in a westerly direction result in longer apparent day length. Movements along the north-south axis affect estimates of apparent noon and midnight, an effect which is magnified at the solstice and minimised at the equinox. Very large movements can produce light observations that do not match expected light observations in any cell along the true flight path taken by the bird, and this was observed in the STSH tracks leading to large errors during migratory periods. One improvement might be to calculate solar zenith angles at each corner of the cell to fully describe when light could be seen within that cell (rather than only at its centre). However, this solution would require a sufficiently large cell size to encompass the daily movement of the animal, potentially increasing discretisation error.

A benefit of larger cell sizes is reduced processing times. Using a 1° × 1° cell on a 30 × 70 grid, we were able to reproduce the 83 day SES track from tag 86372 in just 34.7 s using a 1.9 GHz i5 processor. Decreasing cell size to 1/4° × 1/4° on a 120 × 280 grid increased time to run the track to 31.7 min. The number of cells under consideration at each iteration increases with the proportion of shading assumed, while the number of cells under consideration decreases substantially when SST data are available. A compromise between speed and precision must be decided by the user, but in general, the cell size should be appropriate for the research question, taking into consideration the error inherent in GLS. Cell size will not affect estimates of when
migration or breeding periods occur, nor will small cells improve estimates of migratory routes, but smaller cells may provide better insights into habitat usage. The lower limit of cell size is determined by the amount of movement expected, minimising error by maintaining the assumption that the animal stays within the cell over a day.

Currently, the method is a time invariant HMM (Rabiner, 1989), so that transition probabilities and shading parameters do not change over time. In reality, animals, particularly migratory species, do change movement and behaviour patterns over time (Jonsen, Myers, & James, 2006). A true light record from an STSH, for example, has periods where noise is present and movement is restricted (typically during chick rearing) and periods where noise is almost non-existent and movement is extreme (typically during migration). Figure 6 shows that the daily displacement of both species used in this study do not follow nice, smooth, unimodal distributions. Introducing behavioural states and allowing movement and shading parameters to vary over time should further improve location estimates.

Some of the error observed in this study can be attributed to the methods of comparison. The twilight-free method produces one fix per day, whereas the method used to produce the STSH track (Sumner et al., 2009) produces a fix at each twilight. For simplicity, an average of these positions was taken to produce a single fix for comparison. Putting aside the issue of the average co-ordinate not being on the great circle path between fixes, and very likely not being a location actually visited by the animal, these sorts of comparisons do not account for the movement of the animal. This means that some of the error will depend on the choice of reference point. Even if the estimated locations were perfectly correlated with true locations visited by the animal, we might still estimate large errors. For the positions estimated from the SES data, we decided it was defensible to choose the GPS fix closest to the estimated location on that day for comparison. This gives a fair representation of how close the estimated locations were to known locations visited by the animal over time. In this case the measurement error is still sensitive to animal movement, but inversely proportional to the number of GPS fixes obtained over a day.

The most obvious sources of bias are the calibrated tag threshold and solar zenith angles, as with any solar geolocation method. A less obvious source of bias is the shading parameter. These errors can be understood in terms of day length—the more shading assumed the longer the apparent day length. This pushes location estimates towards the poles in summer and towards the equator in winter. Due to the stochastic nature of the observed data, there is no clear analytical approach to estimating shading (or transition) parameters. Further testing using double-tagging and simulation data is required to develop a set of parameters that yield the most accurate tracks for a range of tagged species. Nevertheless, we have shown that acceptable tracks are estimated even with quite large deviations from the optimal parameters.

The twilight-free method is best suited to analysis of long-range movements of sea-birds, and deep diving species such as elephant seals and penguins. Sea-birds, including albatross and shearwater species, are known to obscure sensors frequently but move across the open ocean. Although other birds are known to obscure sensors frequently, spatially discrete methods may not be suitable if very fine-scale habitat use needs to be identified. The issue of frequent, deep dives attenuating the light is illustrated in the southern elephant seal examples presented in this paper and is a significant challenge for other methods. For other species of marine mammal, where it is relatively easy to mount the GLS tag in such a way that it is not easily obscured, the twilight-free method might be considered as a fast and efficient process for analysing tag data if a single fix per day is sufficiently informative.

There is a clear benefit in double-tagging a subset of animals with satellite and GLS tags, where this is possible, in order to validate GLS tracks (Phillips et al., 2004; Shaffer et al., 2005; Teo et al., 2004). Even where tracks cannot be validated by double-tagging, the standard error of the mean across the movements of a population or species can be reduced by bigger sample sizes, and we hope that the new twilight-free method will increase the number of tracks produced from GLS tags. Previously unusable noisy light records can now be used, and already we have begun work to salvage data already collected but unpublished for lack of a method of analysis. Quantifying where animals are going and when, and monitoring changes in habitat is fundamentally important to conservation and management of many populations and species (Bograd, Block, Costa, & Godley, 2010; Hindell et al., 2011; Phillips, Silk, Croxall, & Afanasyev, 2006; Raymond et al., 2015). Broad scale monitoring requires affordable and reliable tracking technologies. We hope that by effectively increasing the proportion of tags yielding usable data, we can contribute to conservation and monitoring efforts in a meaningful way.

ACKNOWLEDGEMENTS

The original tag data used in the production of the simulated short-tailed shearwater track was collected by N. Bool as part of her doctoral research, and the tag data sampled to produce Figure 1 was provided by J. B. Cleeland from monitoring efforts of albatross populations on Macquarie Island, Tasmania. The authors thank Ms. Bool and Ms. Cleeland for their invaluable support. The corresponding author thanks colleagues in the Hindell Lab, Ecology and Biodiversity Centre, Institute for Marine and Antarctic Studies, University of Tasmania, for support, help, and encouragement.

AUTHORS’ CONTRIBUTIONS

A.D.B., S.J.W. and M.A.H. conceived the idea. A.D.B. and S.J.W. developed the methodology and wrote the R code. C.G. collected SES data. A.D.B., S.J.W.W. and M.A.H. wrote the manuscript. All authors contributed critically to the drafts and gave final approval for publication.

DATA ACCESSIBILITY

Archival tag data and simulation data, including light and SST, and corresponding ground-truth data have been archived at https://github.com/ABindoff/geolocationHMM (https://doi.org/10.5281/zenodo.1065626).
REFERENCES


