# State–space model for light-based tracking of marine animals

**Anders Nielsen and John R. Sibert**

**Abstract:** A coherent model is presented to estimate the most probable track of geographic positions directly from a series of light measurements. The model estimates two geographic positions per day, without reducing the daily light data to two threshold crossing times, its covariance structure is designed to handle high correlations due to for instance local weather conditions, and it can estimate the yearly pattern in latitudinal precision by propagating the data uncertainties through the geolocation process. The model is applied to one mooring study, one GPS drifter buoy study, and numerous simulated cases. The simulations are performed with realistic assumptions about the relationship between solar altitude and light and with realistic uncertainty parameters (all taken from real data). The simulations showed that all model parameters were identifiable, and that all tracks could be reconstructed within 1° or 2° latitude and 0.5° or 1° longitude. The mooring and drifter buoy data showed that the tracks could be reliably estimated, even in cases where the other methods had completely failed.

**Résumé :** Nous présentons un modèle cohérent pour estimer directement le tracé le plus probable des positions géographiques à partir d'une séries de mesures de lumière. Le modèle estime deux positions géographiques par jour, sans réduire les données de lumière journalières à deux périodes de traversée du seuil ; il possède une structure de covariance conçue pour tenir compte des fortes corrélations dues, par exemple, aux conditions climatiques locales ; il peut aussi estimer le patron annuel de précision des latitudes en faisant passer les incertitudes des données dans le processus de géopositionnement. Nous utilisons le modèle sur une étude à bouée fixe, une étude à bouée dérivante munie d'un GPS et plusieurs cas de simulation. Les simulations comportent des présuppositions réalistes concernant la relation entre l'altitude du soleil et la lumière, ainsi que des paramètres d'incertitude réalistes (tous tirés de données réelles). Les simulations montrent que tous les paramètres du modèle sont identifiables et que tous les tracés peuvent être reconstitués avec une précision de 1° ou 2° latitude et de 0,5° ou 1° longitude. Les données obtenues des bouées fixes et dérivantes montrent qu'il est possible d'estimer les tracés de façon fiable, même dans les cas où les autres méthodes ont été complètement incapables de le faire.

[Traduit par la Rédaction]

## Introduction

Archival tags, also referred to as data storage tags, are electronic devices designed to be attached externally or surgically implanted into a live animal to record measurements of the animal's environment. For terrestrial animals and for marine animals that stay at the surface for long periods, satellite-based methods can be used to give a very accurate track. Marine species live in salt water where radio waves cannot penetrate, hence these direct position measurements are impossible. For these species, measurements of light, temperature, and pressure are typically stored in the tags, in the hope that reliable geographic positions (geolocations) can be estimated from these data downloaded from recovered archival tags.

Fisheries research organizations, universities, and governments spend large amounts of funds on development, procurement, and deployment of archival tags. This is not because the technology is fully developed, but simply because there really is no alternative. Directly observing these animals in situ is only possible for very short periods. Archival tags can stay on for years and potentially provide valuable information about after-release survival, spawning areas, habitat preferences, and migration corridors (e.g. Arnold and Dewar 2001; Block 2005).

Algorithms for estimating geolocation based on light and pressure have at least a 20-year history (Smith and Goodman 1986; Musyl et al. 2001; Ekstrom 2004). The algorithms are best described as threshold algorithms. From the depth-corrected time series of light measurements, they determine the time when a certain light threshold is crossed. This threshold is assumed to correspond to a certain solar altitude (i.e. at two times a day the solar altitude is (assumed) known). Solar altitude is measured in degrees above the horizon (negative below). The longitude can be computed from local noon (the midpoint between these two times) and latitude from the time elapsed between these two times (day length). The last part of the algorithms, computing position from two known times and angles, is unproblematic, based in celestial mechanics, and well described in textbooks on astronomical algorithms (Meeus 1998).

The problematic part is associating a threshold with a certain solar altitude, as the time when the daily light measurements crosses a certain threshold is influenced by multiple factors (e.g. local weather conditions). Cloud cover will advance the

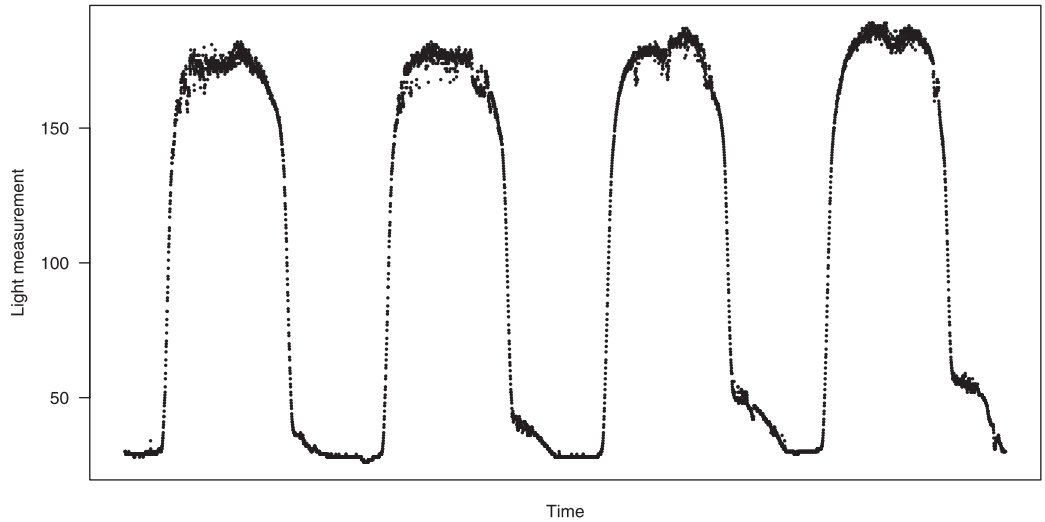
Received 19 December 2005. Accepted 27 June 2006. Published on the NRC Research Press Web site at http://cjfas.nrc.ca/ on 1 August 2007.
J19554

**A. Nielsen[1] and J. R. Sibert.** Pelagic Fisheries Research Program, Joint Institute for Marine and Atmospheric Research, University of Hawai'i at Mānoa, 1000 Pope Road, MSB 312, Honolulu, HI 96822, USA.

[1] Corresponding author (e-mail: anders.nielsen@hawaii.edu).

 doi: 10.1139/F07-064 

**Fig. 1.** Four days of raw light measurements from an archival tag.




time of threshold crossing near sunset and delay it near sunrise, so the latitude determined by the difference will be doubly wrong. Around an equinox, where day lengths are very similar at all latitudes, even small errors in the crossing times will result in huge errors in the estimated latitudes. The longitudinal estimates are generally more accurate, as they are computed from the midpoint of the two daily crossing times. If one crossing time is delayed and the other is equally advanced, the longitudinal estimate is still correct.

A possible improvement is presented by Ekstrom 2004, where it is suggested to fit a template function to narrow-band blue light in the twilight period. In this period, light in these wavelengths may be less sensitive to weather conditions.

Many variants of threshold algorithms are in use, but the details of some algorithms are proprietary, which makes it impossible to present the algorithms here and to thoroughly study them. What can be studied is their performance. Often, tagged marine animals are mistakenly placed hundreds of kilometres from their actual location, which can easily be on dry land (e.g. Welch and Eveson 1999; Musyl et al. 2001; and Sibert et al. 2003). Close to equinoxes, these algorithms frequently fail completely for the latitude coordinate and produce geolocations that are up to 40° of latitude from the true locations (Musyl et al. 2001).

Intuitively the solution is simple. Each position should not be estimated in a knowledge void. The first position is the known release position. The second position must be expected to be in the neighborhood of the first, and so on. A statistically formal way to express this is a state–space model (Harvey 1990). A state–space model works by assuming an underlying movement model, and estimates the track that best matches both the assumed model and the available data. Even a simplistic movement model can be helpful. Each point on the estimated track is a weighted average of the prediction from the assumed movement model and the estimate from the daily light data. The weighting is done according to prediction precision of the assumed movement model and the precision of the data.

The first state–space model for light-based geolocation (Sibert et al. 2003) used the raw light-based daily geolocations as the only data and assumed a simple random walk as the movement model. This model greatly improved the estimated tracks and became widely used (e.g. Musyl et al. 2003, Wilson et al. 2005, and Wilson et al. 2006). The model was extended to also use sea surface temperature (Nielsen et al. 2006), which further improved the estimated track, especially the latitude coordinate in those periods around each equinox where the light-based information is most uncertain.

Using the two step approach, of first estimating raw geolocations by threshold algorithms and then using these estimates as data in a state–space model, is not optimal. The threshold algorithms are not helped by information from previous positions, as each raw geolocation is processed independently. This leads to great uncertainties, especially around equinoxes. Furthermore, the two step approach prevents the state–space model from getting access to all the relevant light data, as the light data are reduced to a (possibly poor) geolocation before it is used in the state–space model. One obvious consequence of this is that any information about the uncertainty of each raw geolocation that might have been in the light data is lost, and the state–space model must re-estimate this information.

This paper presents a coherent state–space model that estimates a track of geolocations directly from light measurements in the tag. All problems associated with threshold algorithms are completely avoided, as they are not a part of the model. This allows all parts of the model movement, covariance structure, and even the functional relation between solar altitude and light measurements to be estimated within the model. The model gives a sound statistical foundation for estimation of the most probable track, and for quantifying its precision.

## Materials and methods

The raw record of an archival tag is a series of measurements taken at times $t_1, t_2, \ldots, t_M$. Typically these times are equidistant, but values can be missing. At these times, three variables are measured: depth, temperature, and light intensity. In the following, only time and light intensity $(t, l)$ are considered.

### Focus on solar event periods

From the raw light record (example in Fig. 1), periods of daytime and nighttime can easily be identified and in between are the periods surrounding sunrise and sunset. Light intensity

is only useful for geolocating if a change in position translates to a change in light intensity. At midnight, even a fairly big change in position (say 5°) will not change the light intensity noticeably — it will still be dark. Similarly at noon, the change in solar altitude resulting from changing position a few degrees will change the light intensity slightly, but local weather conditions, cloud cover, and other uncontrollable features dominates the measured light intensities completely. The periods around solar events (sunrises and sunsets) are the relevant parts for geolocation.

An automatic procedure to select the critical periods around each solar event can be set up in many ways. It is likely not important which procedure is chosen, as long as the resulting intervals include the informative parts of data. The following describes the procedure used in this paper.

First, the raw light record is partitioned into 24 h periods. Within each of these 24 h periods, two observation times are selected such that the mean of the light measurements between the two times ($\widetilde{t}_i$ and $\widetilde{t}_{i+1}$) and the mean of the remaining light measurements times differ the most. In other words:

$$(1) \qquad \widetilde{t}_i, \widetilde{t}_{i+1} = \underset{a,b}{\operatorname{argmax}} \left| \operatorname{mean}(\{l_j : a < t_j \leq b\}) - \operatorname{mean}(\{l_j : t_j \leq a \vee b < t_j\}) \right|$$

The times $\widetilde{t}_1, \widetilde{t}_2, \ldots, \widetilde{t}_{2N}$, where $N$ is the number of days in the light record, are used to select the data intervals to be used in the geolocation model. The interval around each solar event $i$ is chosen asymmetrically to be from 14.4 min (1% of 24 h) towards the nighttime from $\widetilde{t}_i$ to 72 min (5% of 24 h) towards the daytime from $\widetilde{t}_i$. The interval is chosen asymmetrically, because it is not desirable to include points from the dark nighttime period or close to the nighttime period, as moonlight can bias the light measurements in this period. Choosing the length of these intervals (here 86.4 min) is a trade-off. Longer intervals include more data in the analysis, but also increase computation times substantially. Furthermore, the model assumes that horizontal movement within each interval is negligible, and shorter intervals make this assumption more reasonable. Threshold algorithms assume that movement between two solar events is negligible, which is clearly less reasonable.

Observation times within the $i$th interval are denoted $\tau^{(i)} = (\tau_1^{(i)}, \ldots, \tau_{n_i}^{(i)})$, and the corresponding light observations are denoted $l^{(i)} = (l_1^{(i)}, \ldots, l_{n_i}^{(i)})$. Finally, the average observation time within the $i$th interval is denoted $\bar{t}_i$. These average times $\bar{t}_1 \leq \cdots \leq \bar{t}_{2N}$ will be the times where geolocations are computed by the following model (illustrated in Fig. 2).

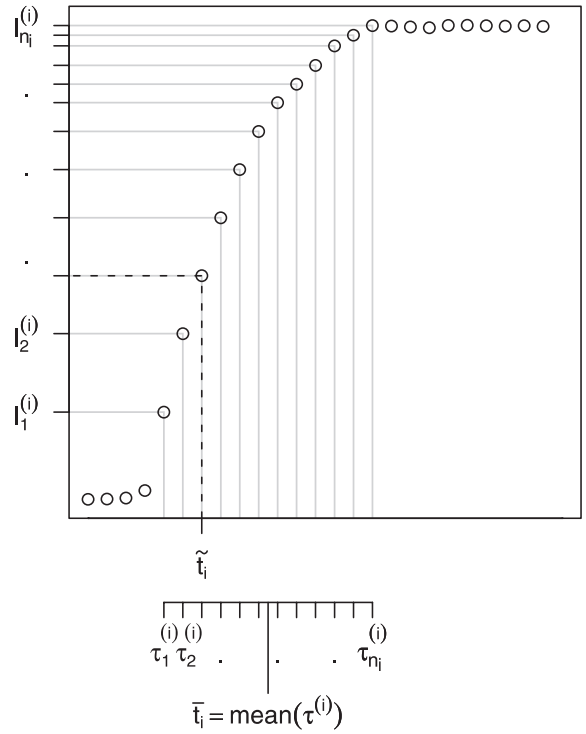
## Model

The model for the observed light measurements along an animal track is a state–space model, where the transition equation describes the movements along the sphere, and the measurement equation describes the predicted light measurements at any given position. For a fixed set of model parameters, the model reconstructs the track where the movements between positions best matches the transition equation, and where the positions yields the best predictions of the actual light measurements.

For the transition equation, a random walk model is assumed:

$$(2) \qquad \alpha_i = \alpha_{i-1} + c_i + \eta_i, \qquad i = 1, \ldots, 2N$$

**Fig. 2.** Illustration of the relationship among $\widetilde{t}_i$, $\tau^{(i)}$, and $\bar{t}_i$ around one solar event.




Here, $\alpha_i$ is a two-dimensional vector containing the coordinates $(\alpha_{i,1}, \alpha_{i,2})$ in nautical miles (1 nautical mile = 1.853 km) along the sphere from a translated origin at time $\bar{t}_i$. $c_i$ is the drift vector describing the deterministic part of the movement, $\eta_i$ is the noise vector describing the random part of the movement, and $N$ is the number of days in the track. The deterministic part of the movement is assumed to be proportional to time $c_i = (u\Delta\bar{t}_i, v\Delta\bar{t}_i)'$. The random part is assumed to be serially uncorrelated and follow a two-dimensional Gaussian distribution with mean vector 0 and covariance matrix $Q_i = 2D\Delta\bar{t}_i I_{2\times 2}$. Here, $D$ is a model parameter expressing the diffusive movement component and $I_{2\times 2}$ is the two-dimensional identity matrix.

The measurement equation of the state–space model describes the expected light measurements at the position $\alpha_i$ at times $\tau^{(i)}$. The calculation of the expected light measurements can be partitioned into three steps. (*i*) Position $\alpha_i$ is transformed into degrees of longitude and latitude by a function $z$ as described in (Sibert et al. 2003). (*ii*) Solar altitudes $\theta$ at the position $z(\alpha_i)$ at times $\tau^{(i)}$ are calculated from standard astronomical algorithms (Meeus 1998). (*iii*) Expected light measurements are calculated from the solar altitudes. The function $\varphi$ describing this relationship is not a known function, and hence it must be estimated within the model.

The function $\varphi$ is represented in the model by a cubic spline function interpolating the points $(\widetilde{\theta}_1, \widetilde{\varphi}_1), \ldots, (\widetilde{\theta}_{n_\varphi}, \widetilde{\varphi}_{n_\varphi})$, where $\widetilde{\theta} \leq \cdots \leq \widetilde{\theta}_{n_\varphi}$ are chosen equidistant over the angles in the data. Notice here that the selection of data intervals around solar events substantially narrows the angle interval where this function is needed. The $\widetilde{\varphi}$s are model parameters. To stabilize the optimization and to reflect our knowledge about the relation between solar altitude and light intensities the $\widetilde{\varphi}$s are restricted to be increasing. Combining steps *i*–*iii* leads to the following

measurement equation:

$$(3) \qquad l^{(i)} = \varphi_{\widetilde{\varphi}}(\theta(z(\alpha_i), \tau^{(i)})) + \varepsilon_i, \qquad i = 1, \ldots, 2N$$

For easier future reference, the entire nonlinear mapping from position to expected light measurement is denoted $\Lambda$, which reduces the measurement equation to

$$(4) \qquad l^{(i)} = \Lambda(\alpha_i, \tau^{(i)}) + \varepsilon_i, \qquad i = 1, \ldots, 2N$$

The measurement error $\varepsilon_i$ is assumed to follow a Gaussian distribution with mean vector 0 and covariance matrix $\Sigma^{(i)}$, where

$$(5) \qquad \Sigma^{(i)}_{j,k} = \begin{cases} \sigma_1^2 + \sigma_2^2 + \sigma_3^2, & \text{if } j = k \\ \sigma_1^2 + \sigma_2^2 \exp(-|\tau_j^{(i)} - \tau_k^{(i)}|/\rho), & \text{if } j \neq k \end{cases}$$

and $j, k = 1, \ldots, n_i$. Here, $\sigma_1$, $\sigma_2$, $\sigma_3$, and $\rho$ are model parameters. Each element $\Sigma^{(i)}_{j,k}$ in this covariance matrix describes the covariance between two light measurements $l_j^{(i)}$ and $l_k^{(i)}$ taken around the same solar event.

The covariance structure in eq. 5 reflects the intuition that two light measurements taken near the same solar event are more similar (correlated) than two taken at separate solar events ($\sigma_1$). If visibility is low one morning, all measurements will be lowered. The covariance structure furthermore allows the correlation between two light measurements near the same solar event to decrease as the time between them increases ($\sigma_2$). Finally, the covariance structure includes a term describing the independent measurement errors ($\sigma_3$).

Now the state–space model is completely defined. All parameters of this model are

$$(6) \qquad \vartheta = (u, v, D, \widetilde{\varphi}_1, \ldots, \widetilde{\varphi}_{n_\varphi}, \sigma_1, \sigma_2, \sigma_3, \rho)$$

The following describes how the model is optimized with respect to these model parameters and how the model is used to reconstruct the most probable track.

## Estimation via the unscented Kalman filter

The basic Kalman filter (Harvey 1990) assumes that both the transition equation and the measurement equation of the state–space model are linear. The extended Kalman filter (Harvey 1990) can handle slight nonlinearities by local first order Taylor approximations of the nonlinear functions in the model. The unscented Kalman filter (Julier et al. 2000) is a more recent sequential estimation technique. Instead of approximating the nonlinear functions, the transformed probability distributions are approximated directly. This approach gives a simpler implementation, not requiring derivatives of the equations in the state–space model; it also allows for higher accuracy — at least corresponding to a second-order Taylor approximation (Julier et al. 2000).

### *The unscented transformation*

Assume that $x$ is a $d$-dimensional random variable following a Gaussian distribution with mean vector $\mu_x$ and covariance matrix $\Sigma_x$. Consider now the distribution of an arbitrary, nonlinear transformation $y = f(x)$

The first-order Taylor approximation, which is used in the extended Kalman filter, states that

$$(7) \qquad y \overset{\text{app}}{\sim} \boldsymbol{N}\left(f(\mu_x), \left(\left.\frac{\mathrm{d}f}{\mathrm{d}x}\right|_{x=\mu_x}\right) \Sigma_x \left(\left.\frac{\mathrm{d}f}{\mathrm{d}x}\right|_{x=\mu_x}\right)'\right)$$

The unscented transformation is a different approximation of the distribution of $y$. The distribution of $x$ is represented by a set $\{x\}$ of $2d$ cleverly selected points:

$$(8) \qquad \{x\} \overset{\text{def}}{=} \mathrm{SP}(\mu_x, \Sigma_x) \overset{\text{def}}{=} \left\{ x \;\middle|\; x = \mu_x \pm \left(\sqrt{d\Sigma_x}\right)_i, \quad i = 1, \ldots, d \right\}$$

where $(\sqrt{d\Sigma_x})_i$ denotes the $i$th column of the Cholesky decomposition of the matrix $d\Sigma_x$.

These so-called sigma points are a simple transformation of the mean and covariance of the distributions they represent, and the notation SP in eq. 8 is the shorthand notation for this transformation.

Each of these sigma points are now transformed by the function $f$ to their corresponding $y$ points:

$$(9) \qquad y^{(j)} = f(x^{(j)}), \qquad j = 1, \ldots, 2d$$

This set of $y$ points representing the transformed distribution is denoted $\{y\}$, and a shorthand for this transformation of $x$ points into $y$ points, which is used in the following, is defined as $\{y\} = f(\{x\})$.

The mean vector (eq. 10) and covariance matrix (eq. 11) of the transformed points are now computed as

$$(10) \qquad \widehat{\mu}_y = \mathrm{mean}(\{y\}) = \frac{1}{2d} \sum_{j=1}^{2d} y^{(j)}$$

$$(11) \qquad \widehat{\Sigma}_y = \mathrm{cov}(\{y\}) = \frac{1}{2d} \sum_{j=1}^{2d} (y^{(j)} - \widehat{\mu}_y)(y^{(j)} - \widehat{\mu}_y)'$$

The approximate distribution of $y$ given by the unscented transformation is the Gaussian distribution with mean vector $\widehat{\mu}_y$ and covariance matrix $\widehat{\Sigma}_y$. With the notation introduced here, the entire unscented transformation can be summarized in

$$(12) \qquad y \overset{\text{app}}{\sim} \boldsymbol{N}\big(\mathrm{mean}(f(\mathrm{SP}(\mu_x, \Sigma_x))), \mathrm{cov}(f(\mathrm{SP}(\mu_x, \Sigma_x)))\big)$$

Notice when comparing eq. 12 with eq. 7 that the approximation in the unscented transformation is simpler to compute, as it does not require derivatives of $f$. All that is needed is to evaluate the nonlinear function in a few carefully selected points.

The unscented transformation is presented here in its simplest form. Other versions include optional scaling of the sigma points. This is introduced to keep the sigma-points closer together when the dimension of $x$ is high. In this application $d = 2$, so these complications are avoided.

### *The unscented Kalman filter*

The unscented Kalman filter is very similar to the extended Kalman filter, except that the unscented transformation is used instead of local Taylor approximations. The following describes the exact unscented Kalman filter equations for this model.

**Step 0:** The filter is started by assuming that the first position $\widehat{\alpha}_0$ in the track is known without error (its covariance is $P_0 = 0_{2\times2}$). This is a reasonable assumption, as the first position is the release position. The initial set of sigma points is computed:

$$(13) \qquad \{\alpha_0\} = \mathrm{SP}(\widehat{\alpha}_0, 0_{2\times2}), \text{ and set } i = 1$$

**Step 1:** From the current set of sigma points and the transition equation (eq. 2), the prediction of the next position is calculated, along with its covariance and its sigma point representation:

$$\widehat{\alpha}_{i|i-1} = \text{mean}(\{\alpha_{i-1}\} + c_i) \tag{14}$$

$$P_{i|i-1} = \text{cov}(\{\alpha_{i-1}\} + c_i) + Q_i \tag{15}$$

$$\{\alpha_{i|i-1}\} = \text{SP}(\widehat{\alpha}_{i|i-1}, P_{i|i-1}) \tag{16}$$

**Step 2:** The predicted light level vectors (eq. 17) corresponding to the sigma points are calculated via eq. 4, along with their mean vector, covariance, cross covariance, and a matrix known as the Kalman gain matrix, which is used when updating the predicted position with the actually observed light vector:

$$\{l^{(i)}\} = \Lambda(\{\alpha_{i|i-1}\}, \tau^{(i)}) \tag{17}$$

$$\widehat{l}^{(i)} = \text{mean}(\{l^{(i)}\}) \tag{18}$$

$$F_i = \text{cov}(\{l^{(i)}\}) + \Sigma^{(i)} \tag{19}$$

$$G_i = \text{cov}(\{\alpha_{i|i-1}\}, \{l^{(i)}\}) \tag{20}$$

$$K_i = G_i F_i^{-1} \tag{21}$$

**Step 3:** Finally, the difference between the predicted light vector and the observed is calculated, the Kalman update equations are applied, and an updated set of sigma points are calculated:

$$w_i = l^{(i)} - \widehat{l}^{(i)} \tag{22}$$

$$\widehat{\alpha}_i = \widehat{\alpha}_{i|i-1} + K_i w_i \tag{23}$$

$$P_i = P_{i|i-1} - K_i F_i K_i' \tag{24}$$

$$\{\alpha_i\} = \text{SP}(\widehat{\alpha}_i, P_i) \tag{25}$$

**Step 4:** If $i < 2N$, increase $i$ by one and go to step 1.

***Negative log-likelihood function***

The maximum likelihood principle is used to find the estimates of all model parameters (eq. 6). With the definition that $w_0 = l^{(0)} - \Lambda(\widehat{\alpha}_0, \tau^{(0)})$ and $F_0 = \Sigma^{(0)}$, the negative log-likelihood is given by

$$\ell(\vartheta) = -\sum_{i=0}^{2N} \log \phi(w_i, 0, F_i) \tag{26}$$

Here, $\phi(w, 0, F)$ is the density of the Gaussian distribution with mean vector 0 and covariance matrix $F$ evaluated in $w$.

***Last point known without error***

Knowing the last position in the track corresponds to extending the last observational vector $l^{(2N)}$ with the known two-dimensional position to $l^{(2N)} = (l^{(2N)}, \alpha_{2N})'$ (here expressed in nautical miles coordinates). The model extension to accommodate this extra information is straightforward. $\Lambda$ is extended to also return the predicted position, and the observational covariance matrix $\Sigma^{(2N)}$ is extended by two zero-rows and two zero-columns to express that the position is known without error. Everything else remains unchanged.

***Most probable track***

The unscented Kalman filter and the maximum likelihood principle supply estimates of the model parameters and the predicted track. A point on the predicted track at any given time point is calculated using all observations available at that time, that is, $\widehat{\alpha}_i \approx E(\alpha_i | l^{(1)}, \ldots, l^{(i)})$. Once the entire set of light observations are available, it is possible to obtain estimates with smaller variance. What is here denoted "the most probable track" is calculated using all observations, after the parameters have been estimated. A point on the most probable track is $\widehat{\alpha}_{i|2N} \approx E(\alpha_i | l^{(1)}, \ldots, l^{(2N)})$.

The actual computation of the most probable track is done in a single backwards sweep of the predicted track and its covariance. The last point of the most probable track is identical to the last point of the predicted track, as all observations were available to the predicted track at the final point. The last point of the most probable track is $\widehat{\alpha}_{2N|2N} = \widehat{\alpha}_{2N}$, and its covariance is $P_{2N|2N} = P_{2N}$. The following equations are used to recursively compute the previous points of the most probable track:

$$\widehat{\alpha}_{i|2N} = \widehat{\alpha}_i + P_i^\star(\widehat{\alpha}_{i+1|2N} - \widehat{\alpha}_i - c_{i-1}) \tag{27}$$

$$P_{i|2N} = P_i + P_i^\star(P_{i+1|2N} - P_{i+1|i})P_i^{\star\prime} \tag{28}$$

$$\text{where} \quad P_i^\star = P_i P_{i+1|i}^{-1}$$

In general, it is not possible to use these simple formulas with the unscented Kalman filter, but the transition equation (eq. 2) of this model is linear, so here it is a valid approximation. A general approach is described in Wan and van der Merwe (2001). In textbooks on the Kalman filter (Harvey 1990), this technique is known as smoothing, as the resulting track most often is smoother than the predicted track.

**Simulation of realistic data**

Simulating artificial, but realistic, data from the actual model, is an important tool to investigate most important properties of the model. The advantage of using simulated data is that they are cheap, the true tracks and model parameters are always known, and they can be tailored to focus on specific aspects of the model.

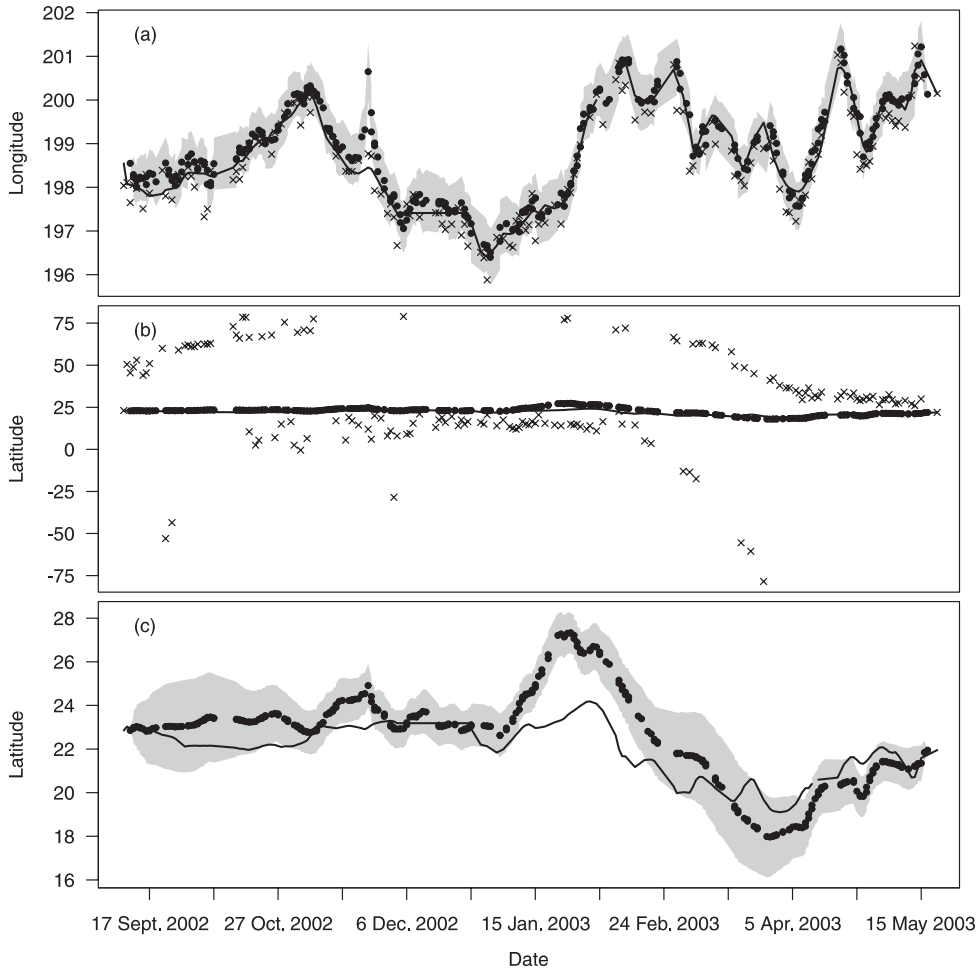
From a set of known model parameters $\vartheta$ and a known first position $\alpha_0 = z^{-1}(y_0)$, the following steps are used to simulate data.

($i$) The underlying true movements are simulated forward in time from the first position via the movement parameters $(u, v, D)$ and the transition equation (eq. 2). The times used $(t_1, t_2, \ldots, t_M)$ should mimic light sampling rates from a real archival tag. Depending on tag type, they could be 20 s or 5 min apart.

($ii$) At each point in the true simulated track, the predicted light intensity is calculated via the function $\Lambda$ in eq. 4 and the parameters $\widetilde{\varphi}_1, \ldots, \widetilde{\varphi}_{n_\varphi}$ describing the function $\varphi$ relating the solar altitude to the tag-measured light intensity. Real estimates of model parameters from an actual tag are used here to make the simulated data as realistic as possible.

($iii$) The uncorrelated part of the observation noise, $r_t \sim \boldsymbol{N}(0, \sigma_3^2)$, $t_1, \ldots, t_M$, is now added to each of these light predictions.

**Fig. 3.** Points on the most probable track from the light-based state–space model (circles), along with raw geolocations from the tag manufacturer's threshold algorithm (crosses) and the true track from the satellite-based method (solid line). The shaded areas are 95% confidence regions estimated by the light-based state–space model. Data indicate longitude coordinate (*a*); latitude coordinate including raw geolocations (*b*); and latitude zoomed to scale of most probable and true track (*c*).




(*iv*) The scanning algorithm, previously described, is applied to select the solar event times and the surrounding times and simulated light intensities.

(*v*) Finally, the remaining part of the observation noise is added. To the simulated light intensities surrounding the $i$th solar event is added a vector $e^{(i)} \sim N(0, S^{(i)}), i = 1, \ldots, 2N$, where $S^{(i)}$ is equal to $\Sigma^{(i)}$ described in eq. 5, but with $\sigma_3^2 = 0$.

The resulting data are a set of time vectors and matching simulated light observation vectors $(\tau^{(i)}, l^{(i)})_{i=1,\ldots,2N}$. These artificial observations will follow the model, except for one minor detail. The model assumes no movement within the time intervals surrounding each solar event, but the simulation movement continues in these intervals. This is done to better mimic real data, and to investigate the consequences of this model assumption.

## Results

A Wildlife Computers pop-up archival transmitting tag (version 2) was deployed on a drifter buoy near Hawaii in September 2002 and set to pop up after 9 months. The software from the manufacturer was used to estimate threshold-based geolocations. The resulting latitude reconstruction was extremely inaccurate (Fig. 3). The data from this type of tag consist of $\sim$ 12 light measurements around each solar event, taken $\sim$ 8 min apart. These raw light measurements were used in the described light-based state–space model, without first applying the scanning described. The estimated most probable track, especially the latitude coordinate, was greatly improved (Fig. 3). The estimated 95% confidence region around the most probable track is not perfect, but it does seem to have approximately the correct size. The tag manufacturer's algorithm for selecting the light measurements to be included in the data does not seem as robust as the one described in this paper, so even better results can be expected if the observation window was selected using that instead.

An important pattern is clearly identified in the standard deviation of the most probable track (Fig. 4), The standard deviation of the longitude is fairly constant throughout the year, but the latitude standard deviation is increasing (doubling) around the two equinoxes. This pattern is not assumed in the model, as was necessary in Sibert et al. (2003) and Nielsen et al. (2006), but is a simple consequence of propagating the data uncertainties

**Fig. 4.** Estimated standard deviation (SD) of the geolocations from the light-based state–space model. Vertical broken lines indicate time of equinox. Data indicate longitude coordinate (*a*) and latitude coordinate (*b*).

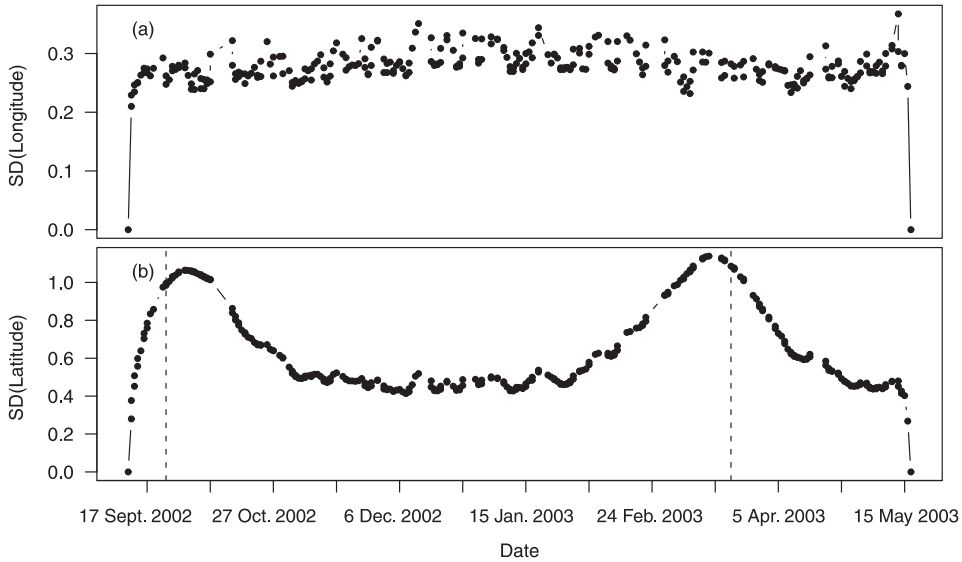



**Fig. 5.** Points on the most probable track from the light-based state–space model (circles here are so close that they appear as a thick solid line), along with raw geolocations from the tag manufacturer's threshold algorithm (crosses) and the true track from satellite-based method (solid line). The shaded areas are 95% confidence regions estimated by the light-based state–space model. Data indicate longitude coordinate (*a*); latitude coordinate including raw geolocations (*b*); and latitude zoomed to scale of most probable and true track (*c*).

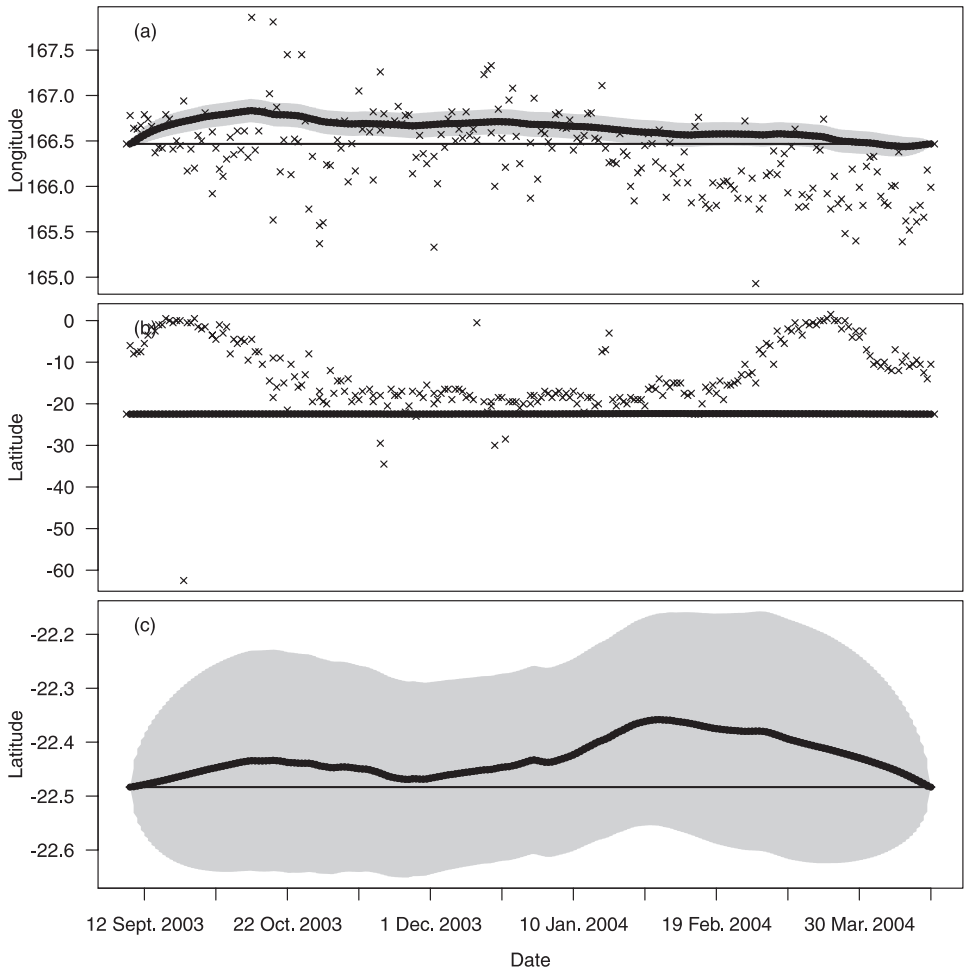
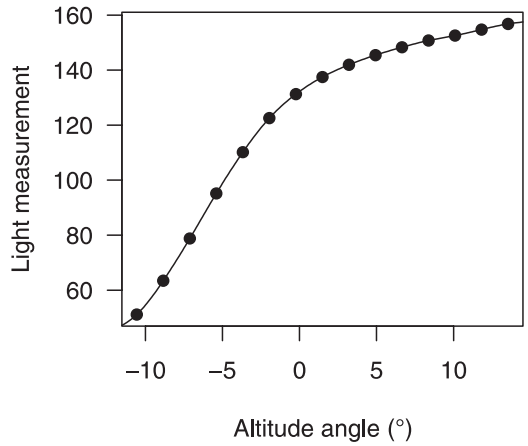


**Fig. 6.** The estimated function $\varphi$ describing the relationship between solar altitude angle and the expected light measurement in the tag. The circles are the estimated support points corresponding to the model parameters $\widetilde{\varphi}_1, \ldots, \widetilde{\varphi}_{n_\varphi}$.




through geolocation process.

A Wildlife Computers archival tag (Mk9, version 1.03) was attached to a mooring near New Caledonia in September 2003, and recorded data for close to 8 months. The software from the manufacturer was used to estimate threshold-based geolocations. The reconstructed latitude was extremely inaccurate (Fig. 5). The raw data contains a light reading every minute. The most probable track from the light-based state–space model is significantly biased for the longitude coordinate in the beginning of the track, whereas the threshold-based longitudes are similarly biased near the end of the track. This bias is likely a result of the internal clock in the tag being slightly biased or drifting. The latitude coordinate of the most probable track is within 0.2° of the true position at all times, and the latitude confidence region covers the true position. Even including the longitude bias, the most probable track is within 0.5° of the true position at all times.

The function $\varphi$ describing the relationship between the solar altitude and the expected light measurement is an important part of the model. The function is estimated within the model for the interval covering the selected light measurements. The function appears to reliably estimated, as the estimated support points are in close agreement with a monotonic increasing curve (Fig. 6). The function $\varphi$ estimated from this mooring study is used as the basis for the following simulations.

To test the model with realistic movement parameters, two scenarios were simulated: first, with a moderate diffusion rate of $D = 100$ Nm$^2$·day$^{-1}$, which is in the range of sea turtle movements (Y. Swimmer, NOAA Fisheries, Pacific Island Fisheries Science Center, 2570 Dole Street, Honolulu, HI 96822, Yonat.Swimmer@noaa.gov, personal communication); second, with a high diffusion rate of $D = 500$ Nm$^2$·day$^{-1}$, which is in the range of bigeye tuna (*Thunnus obesus*) movement (Sibert et al. 2003). Three tracks, at three different release latitudes, were simulated for each scenario. The actual estimated values from the mooring data were used for the functional relationship between the solar altitude and light (Fig. 6) and for the parameters determining the variance structure ($\sigma_1^2 = 83$, $\sigma_2^2 = 9.9$, $\sigma_3^2 = 2.6$, and $\rho = 3.2\%$). This makes the conclusions based on these simulations regarding precision of the most probable track as realistic as possible.

**Table 1.** True parameter values, means, and standard deviations (SD) of estimates from 100 simulated tracks.

| Parameter name | True value | Mean of estimates | SD of estimates |
|---|---|---|---|
| $u$ | 0.000 | –0.138 | 1.366 |
| $v$ | 0.000 | 0.123 | 1.310 |
| $D$ | 300.000 | 295.021 | 34.916 |
| $\rho$ | 0.032 | 0.034 | 0.006 |
| $\sigma_1^2$ | 83.000 | 81.829 | 5.041 |
| $\sigma_2^2$ | 9.900 | 10.424 | 1.281 |
| $\sigma_3^2$ | 2.600 | 2.606 | 0.075 |
| $\varphi_{-5}$ | 98.843 | 98.793 | 0.550 |
| $\varphi_0$ | 132.255 | 132.242 | 0.407 |
| $\varphi_5$ | 145.571 | 145.487 | 0.353 |

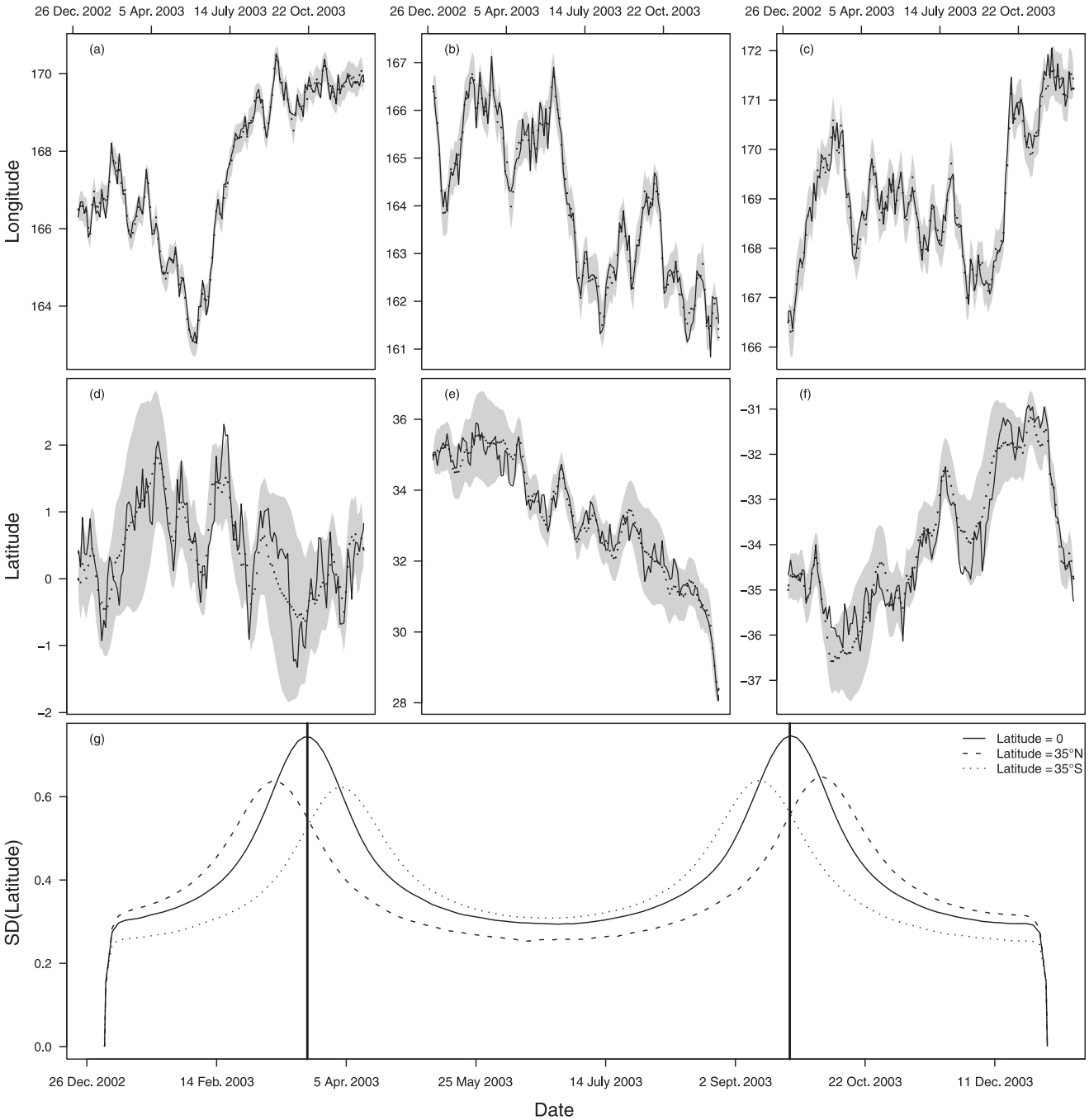
The moderately moving tracks were all estimated accurately (Fig. 7). The longitude coordinate is within 0.5° of the true track, and the latitude coordinate is within 1° of the true track. The estimated 95% confidence region appears to have the correct coverage. The standard deviation of the most probable latitude show an interesting pattern. Around each equinox, the standard deviation increases, but not in the same way for all three tracks. The track near the equator (panels *a* and *d*) has the highest latitude standard deviation exactly at the equinox, but for the northern track (panels *b* and *e*) and the southern track (panels *c* and *f*), the time of the highest latitude standard deviation is shifted towards the winter side. Further, it is noted that the standard deviation is higher for the track near equator, where annual variation in day length is least.

The fast-moving tracks were also estimated accurately (Fig. 8). The longitude coordinate is within a degree of the true track, and the latitude coordinate is within 2° of the true track, but most often closer. The estimated 95% confidence region appears to have the correct coverage. The standard deviation of the most probable latitude standard deviation shows the same pattern as for the moderately moving tracks, except that the standard deviation is higher for the fast-moving tracks. This is expected, as larger diffusion $D$ leads to larger uncertainty in the prediction from the underlying movement model.

To verify that all model parameters in this model are identifiable, 100 tracks were simulated from known parameters, and the light-based state–space model was applied to each track. The parameters representing the relationship between solar altitude and light $\widetilde{\varphi}_1, \ldots, \widetilde{\varphi}_{n_\varphi}$ are chosen dynamically to cover the range (of angles) of the selected light measurements, and hence they may represent the light value at slightly different angles for each simulated track. To be able to verify the estimated relationship between between solar altitude and light three fixed angles are selected: –5°, 0°, and 5°. For each simulated track, the estimated light levels ($\varphi_{-5}$, $\varphi_0$, and $\varphi_5$) at these three angles are evaluated.

The simulation study shows that all model parameters are identifiable. Including the relationship between solar altitude and light (Table 1), the movement parameters ($u$, $v$, and $D$) show no significant bias. The relationship between solar altitude and light is estimated unbiased at –5° and 0°, but these 100 simulations indicate a slightly biased estimate (of 0.084 light units) at 5°, but this is likely a coincidence occurring because of the relative low number of simulations.

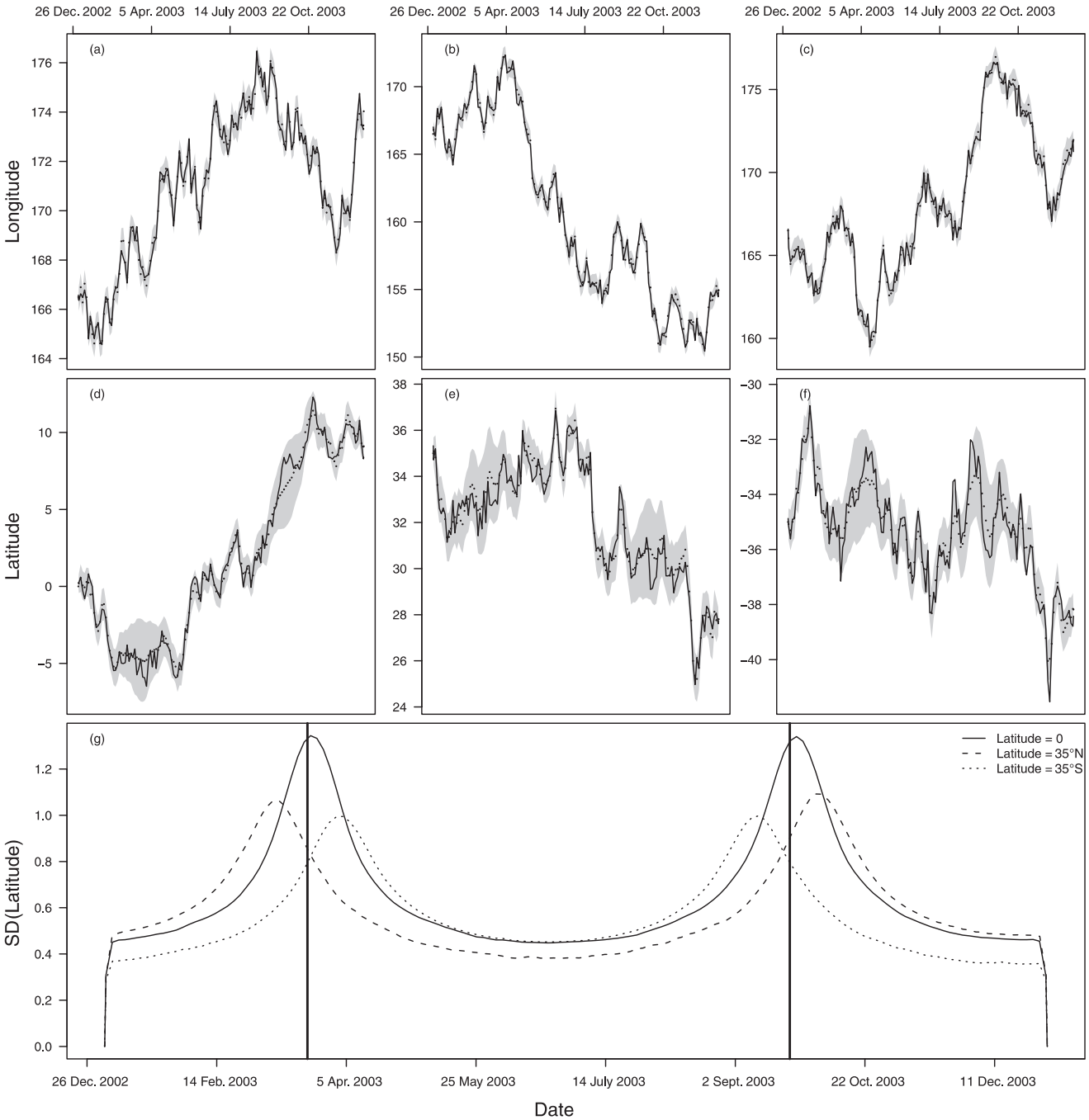
**Fig. 7.** Longitude (*a*, *b*, and *c*) and latitude (*d*, *e*, and *f*) of three moderately moving simulated tracks: released at equator (*a* and *d*), released at 35°N (*b* and *e*), and released at 35°S (*c* and *f*). Each frame shows the true simulated track (solid line), the most probable track from the light-based state–space model (dots), and the 95% point-wise confidence region (shaded area). Finally, the estimated standard deviation (SD) of the most probable latitude is shown (*g*).




Of the estimated parameters in the variance structure ($\sigma_1$, $\sigma_2$, $\sigma_3$, and $\rho$), three are slightly biased. The parameter describing the variation in visibility between different solar events ($\sigma_1$) is negatively biased, but the parameters describing the correlation between close observations within the same solar event ($\sigma_2$ and $\rho$) are positively biased. The estimates are trading a small part of the overall correlation between light measurements from the same solar event to higher correlation between neighboring points. This is a consequence of the model assumption of no movement within the observational window around each solar event. The simulated tracks, and likely also real tracks, continue to move around solar event times, which gives this higher correlation between neighboring points. The actual value of these parameters are rarely used, so this small bias is unproblematic.

In all previous simulations (Figs. 7 and 8; Table 1), the sampling rate has been set to 300 measurements per day, which is one every 4.8 min. The observation window refers to the interval around each solar event where the light measurements are

**Fig. 8.** Longitude (*a*, *b*, and *c*) and latitude (*d*, *e*, and *f*) of three fast-moving simulated tracks: released at the equator (*a* and *d*), released at 35°N (*b* and *e*), and released at 35°S (*c* and *f*). Each frame show the true simulated track (solid line), the most probable track from the light-based state–space model (dots), and the 95% point-wise confidence region (shaded area). Finally, the estimated standard deviation (SD) of the most probable latitude is shown (*g*).

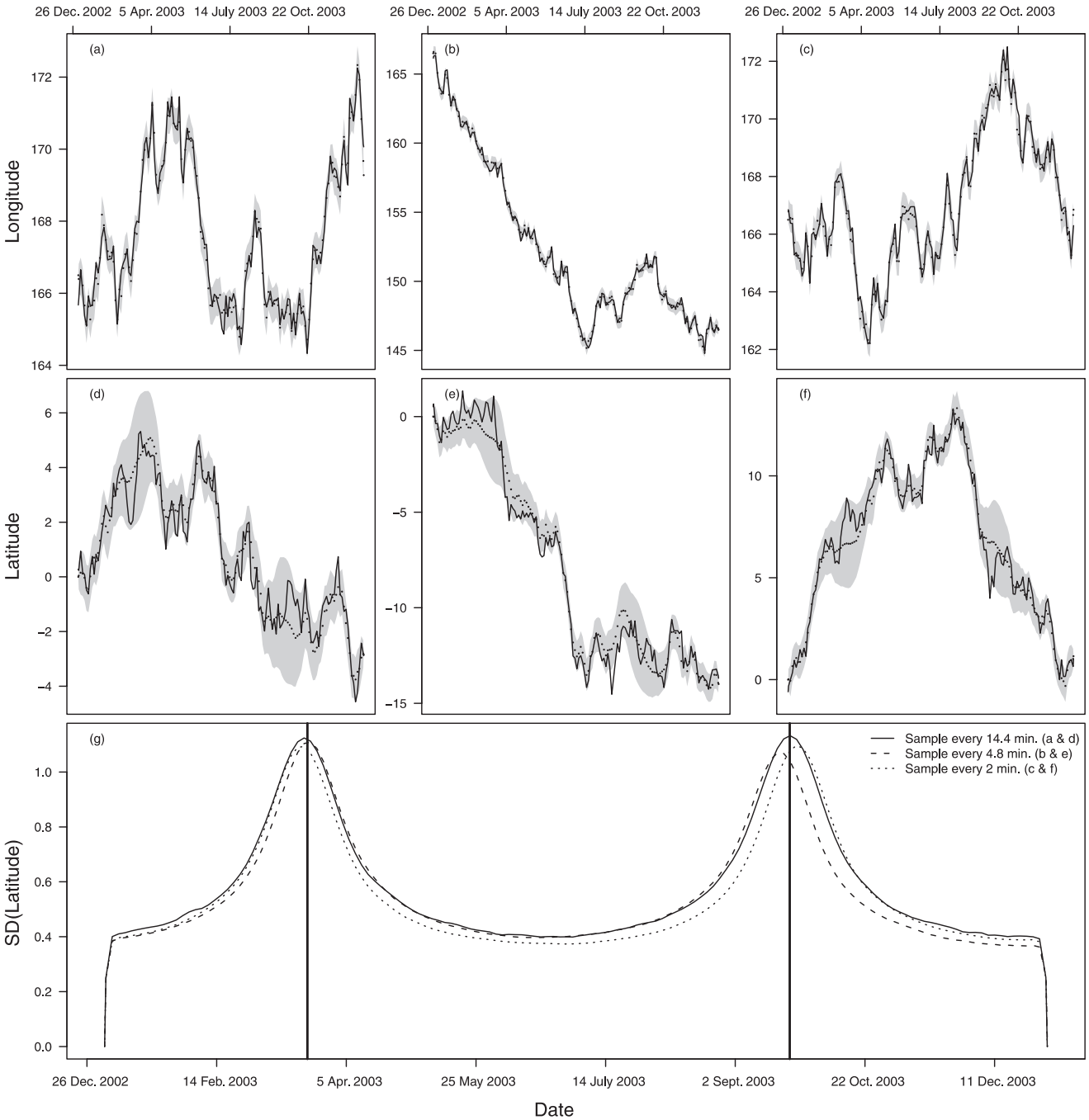



actually used by the model. The observation window, consisting of all observations from the solar event time and 72 min towards the daytime, and all observations from the solar event and 14.2 min towards the nighttime has previously been used.

The sampling rate was varied from every 2 min to every 14.2 min. In all cases, the model was able to follow the true track (Fig. 9). The standard deviation of the most probable latitude was higher at the sparse sampling rates, but the seasonal differences are much more dominant (Fig. 9).

The observation window was tested by using fewer of the observations on the day side of the solar event, or fewer of the observations on the night side of the solar event, or both. In all cases, the model was still able to follow the true track (Fig. 10). The standard deviation of the most probable latitude was higher in the case where only 18 min towards the day and 3.6 min towards the night was used (one-quarter of the original window), but all other tested window sizes gave surprisingly similar latitude variances.

**Fig. 9.** Longitude (*a*, *b*, and *c*) and latitude (*d*, *e*, and *f*) of three simulated tracks: light sampling every 2 min (*a* and *d*), every 4.8 min (*b* and *e*), and every 14.2 min (*c* and *f*). Each frame shows the true simulated track (solid line), the most probable track from the light-based state–space model (dots), and the 95% point-wise confidence region (shaded area). Finally, the estimated standard deviation (SD) of the most probable latitude is shown (*g*).

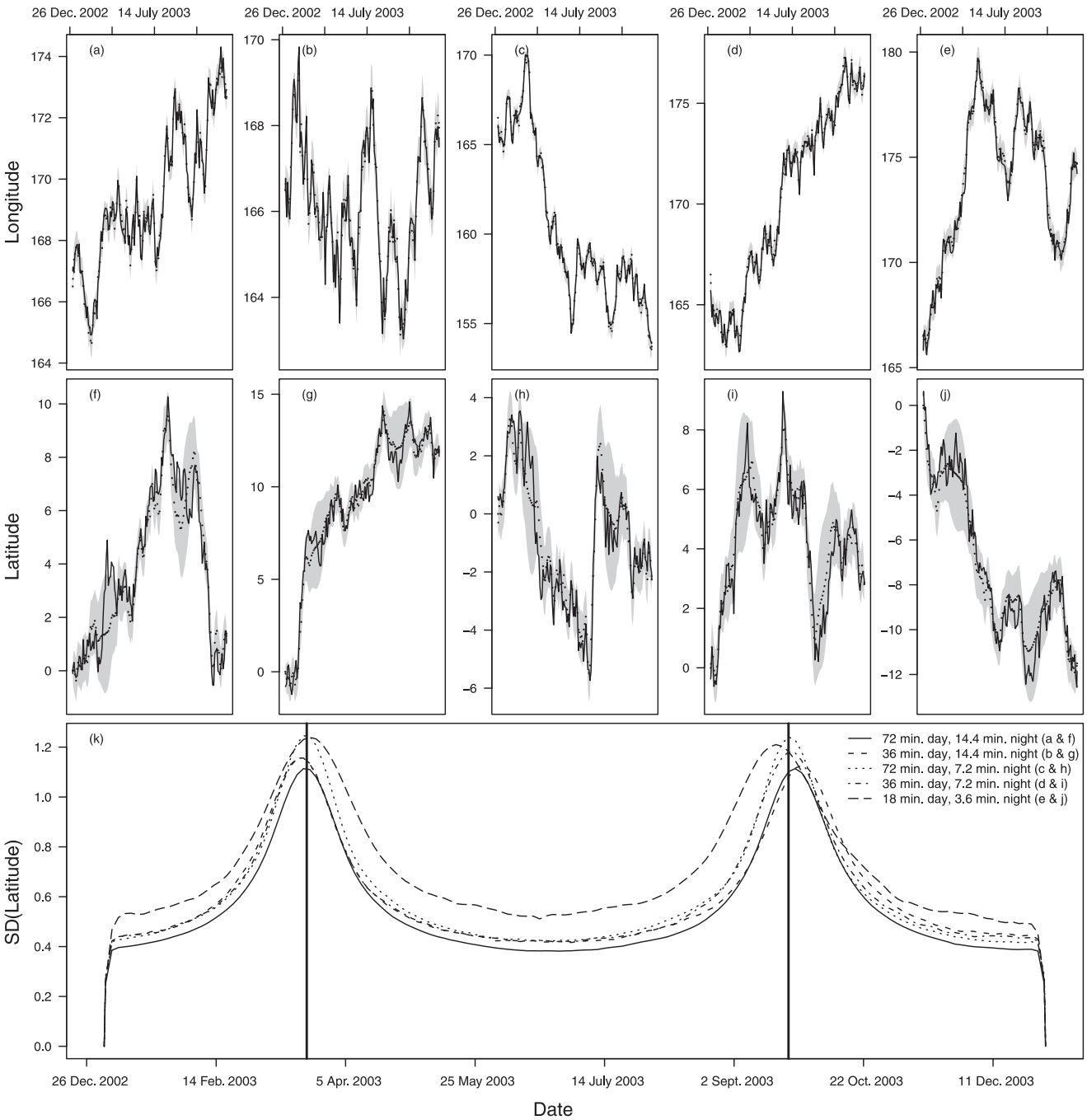



## Discussion

The model presented in this paper is completely different from previous algorithms used in this field. It is unique in combining all steps from the raw light measurements to the estimated most probable track in one coherent model. It cannot be classified as a threshold model, as it does not reduce the light measurements at each solar event to a time where a certain solar altitude is assumed known. It cannot be classified as a template-matching method (as described in Ekstrom 2004) either, as it does not assume to know the relationship between solar altitude and light. Instead it estimates this relationship within the model.

The model presented here is also unique in offering a way to deal with correlated light measurements. It is a fact that light measurements are serially correlated. Assuming they are independent will lead to false confidence statements. The value of each light measurement is influenced by many factors, which

**Fig. 10.** Longitude (*a*, *b*, *c*, *d*, and *e*) and latitude (*f*, *g*, *h*, *i*, and *j*) of five simulated tracks with varying observation window: 72 min towards the day and 14.4 min towards the night (*a* and *f*), 36 min towards the day and 14.4 min towards the night (*b* and *g*), 72 min towards the day and 7.2 min towards the night (*c* and *h*), 36 min towards the day and 7.2 min towards the night (*d* and *i*), and 18 min towards the day and 3.6 min towards the night (*e* and *j*). Each frame shows the true simulated track (solid line), the most probable track from the light-based state–space model (dots), and the 95% point-wise confidence region (shaded area). Finally, the estimated standard deviation (SD) of the most probable latitude is shown (*g*).




are not all included in the model, the most obvious one being local weather conditions. If there is a heavy cloud cover one morning, but clear blue sky the next, the measurements from the latter will all be higher even though all other conditions are equal. The local weather conditions, or other factors influencing the light measurements, can change gradually, so the correlation structure further allows the correlation between two light measurements near the same solar event to depend on the elapsed time between them.

Another unique feature of the model presented here is that geolocations are estimated twice a day. Threshold algorithms need to know the solar altitude angle at two times to produce one geolocation, and obtain this from two solar events. Getting twice as many geolocations is naturally preferable, but the

underlying difference in model assumptions is even more important. Threshold algorithms assume that the movement of the tagged animal between two solar events can be neglected. The model presented here only assumes that movement in the approximately 90 min observational window around each solar event can be neglected. Furthermore, if there are substantial movements within the observational windows, the model will interpret that as additional serial correlation, which will correctly propagate into wider confidence regions.

The ability to estimate the yearly variation in the latitudinal standard deviation is an additional benefit of, and unique to, the model presented here. In Sibert et al. (2003) and Nielsen et al. (2006), an ad hoc parametric model had to be introduced to compensate for this pattern in threshold-based geolocations. It was difficult to verify that the assumed parametric structure was a valid approximation. The model presented here includes no assumptions about the yearly variation in latitudinal standard deviation, but the pattern clearly emerges as a simple consequence of propagating the data uncertainties through the geolocation process.

The two real data cases presented here show that the model presented is able to produce geolocations even in cases where the threshold-based latitudinal estimates are extremely inaccurate. The estimated most probable tracks are not perfect, but they are far superior to the threshold-based geolocations.

All model parameters can be identified from data. The estimated most probable track is able to follow the true track when realistic assumptions about movement pattern and observation uncertainty are made.

The model is surprisingly insensitive to varying the sampling rate of the light measurements. This implies that it is actually possible to get reliable geolocations from a small number (say 10) of correctly selected points around each solar event. This result is potentially very useful, as tags that transmit their data to satellites can only transfer a limited amount of data.

The fact that the model parameters can be identified correctly from data, and that the most probable tracks follow the true tracks, implies that approximation used in the unscented Kalman filter is sufficiently accurate for the nonlinearities in this model. This is fortunate, as the more accurate alternatives (simulation-based filters, e.g., Doucet et al. 2001) are far more computationally demanding.

Ideally, the light that is reflected by the moon should be part of the model, as it clearly influences the light measurements in the periods surrounding each full moon. The astronomical algorithms describing moon altitude and fraction of moon disc illuminated are straightforward to implement (Meeus 1998). Adding the two light sources (sun and moon) is the problematic part. Light measurements from the tag are processed by the inner workings of the tag. This processing can be viewed as a nonlinear transformation of the raw light intensities, and the details of this transformation are not easily obtainable, as the inner workings of some tags are not well documented or are simply considered proprietary. The asymmetric observation window chosen here avoids most of the complications with the moon, as it focus on the part of the light record where moonlight is negligible, but including the moon could potentially further improve this model.

The problems associated with standardizing light measurements taken at different depths, also referred to as depth-correcting, has been avoided in this paper, because the analyzed data were obtained at constant depths. The light measurements can be depth-corrected by the tag manufacturers before they are used in the model, or the model can be extended by a parametric model for the relationship between depth and light. It is expected that depth-correcting will introduce more uncertainty in the light measurements and consequently in the estimated tracks.

The suggestions to use narrow-band blue light instead (Ekstrom 2004) or to use cosine-collectors on the tags (Qayum et al. 2006) are interesting. If these light measurements are less sensitive to weather conditions or better capture light at depths, it will reduce the uncertainties and further improve the most probable track.

It has been shown that using sea surface temperature can improve geolocations (Teo et al. 2004; Nielsen et al. 2006). The model presented here can be extended in exactly the same way. However, instead of using sea surface temperature as a second patch (the first one being the Kalman filter) to a light-based algorithm that is giving extremely inaccurate geolocations, here it would only be used to further improve already reliable light-based geolocations.

## Acknowledgements

The work described in this paper was sponsored by the University of Hawai'i Pelagic Fisheries Research Program under cooperative agreement No. NA17RJ1230 from the National Oceanic and Atmospheric Administration. We thank Kevin Wong (NOAA – Pacific Islands Fisheries Science Center) and Phil White, crew, and officers of the NOAA research ship *Townsend Cromwell* for deploying the drifter buoy and pop-up satellite archival tag, and Mike Musyl (NOAA – National Marine Fisheries Service) for providing the data. We thank Bruno Leroy from the Oceanic Fisheries Programme of the Secretariat of the Pacific Community for providing data from the New Caledonia mooring.

## References

Arnold, G., and Dewar, H. 2001. Electronic tags in marine fisheries research: a 30-year perspective. *In* Electronic tagging and tracking in marine fisheries reviews: methods and technologies in fish biology and fisheries. *Edited by* J.R. Sibert and J.L. Nielsen. Kluwer Academic Press, Dordrecht. pp. 7–64.

Block, B.A. 2005. Physiological ecology in the 21st century: advancements in biologging science. Integr. Comp. Biol. **45**: 305–320.

Doucet, A., de Freitas, N., and Gordon, N. 2001. Sequential Monte Carlo methods in practice. Springer-Verlag, New York.

Ekstrom, P.A. 2004. An advance in geolocation by light. Mem. Natl Inst. Polar Res. Spec. Issue, **58**: 210–226.

Harvey, A.C. 1990. Forecasting, structural time series models and the Kalman filter. Cambridge University Press, Cambridge.

Julier, S., Uhlmann, J., and Durrant-Whyte, H.F. 2000. A new method for the nonlinear transformation of means and covariances in filters and estimators. IEEE Trans. Autom. Control, **45**(3): 477–482.

Meeus, J. 1998. Astronomical algorithms. Willmannn–Bell. Richmond, Va.

Musyl, M.K., Brill, R.W., Boggs, C.H., Curran, D.S., Kazama, T.K., and Seki, M.P. 2003. Vertical movements of bigeye tuna (*Thunnus obesus*) associated with islands, buoys, and seamounts near the main

Hawaiian Islands from archival tagging data. Fish. Oceanogr. **12**: 152–169.

Musyl, M.K., Brill, R.W., Curran, D.S., Gunn, J.S., Hartog, J.R., Hill, R.D., Welch, D.W., Eveson, J. P., Boggs, C. H., and Brainard, R.E. 2001. Ability of archival tags to provide estimates of geographical position based on light intensity. *In* Electronic tagging and tracking in marine fisheries reviews: methods and technologies in fish biology and fisheries. *Edited by* J.R. Sibert and J.L. Nielsen. Kluwer Academic Press, Dordrecht. pp. 343–368.

Nielsen, A., Bigelow, K.A., Musyl, M.K., and Sibert, J.R. 2006. Improving light-based geolocation by including sea surface temperature. Fish. Oceanogr. **15**: 314–325.

Qayum, H.A., Klimley, A.P., Newman, R., and Richert, J.E. 2006. Broad-band versus narrow-band irradiance for estimating latitude by archival tags. Mar. Biol. **151**: 467–481. doi:10.1007/s00227-006-0514-y.

Sibert, J., Musyl, M.K., and Brill, R.W. 2003. Horizontal movements of bigeye tuna (*Thunnus obesus*) near Hawaii determined by Kalman filter analysis of archival tagging data. Fish. Oceanogr. **12**: 141–151.

Smith, P., and Goodman, D. 1986. Determining fish movement from an "archival" tag: precision of geographical positions made from a time series of swimming temperature and depth. NOAA Tech. Memo. NOAA-TM-NMFS-SWFC-60.

Teo, S.L.H., Boustany, A., Blackwell, S., Walli, A., Weng, K.C., Block, B.A. 2004. Validation of geolocation estimates based on light level and sea surface temperature from electronic tags. Mar. Ecol. Prog. Ser. **283**: 81–98.

Wan, E.A., and van der Merwe, R. 2001. The unscented Kalman filter. *In* Kalman filtering and neural networks. *Edited by* S. Haykin. Wiley Publishing, New York. pp. 221–280.

Welch, D.W., and Eveson, J.P. 1999. An assessment of light-based geoposition estimates from archival tags. Can. J. Fish. Aquat. Sci. **56**: 1317–1327.

Wilson, S.G., Lutcavage, M.E., Brill, R.W., Genovese, M.P., Cooper, A.B., and Everly. A.W. 2005. Movements of bluefin tuna (*Thunnus thynnus*) in the northwestern Atlantic Ocean recorded by pop-up satellite archival tags. Mar. Biol. **146**: 409–423.

Wilson, S.G., Polovina, J.J., Stewart, B.S., and Meekan, M.G. 2006. Movements of whale sharks (*Rhincodon typus*) tagged at Ningaloo Reef, Western Australia. Mar. Biol. **148**: 1157–1166.