

# Improving light-based geolocation by including sea surface temperature

ANDERS NIELSEN,<sup>1,\*</sup> KEITH A. BIGELOW,<sup>2</sup>  
MICHAEL K. MUSYL<sup>1</sup> AND JOHN R. SIBERT<sup>1</sup>

<sup>1</sup>*Pelagic Fisheries Research Program, Joint Institute for Marine and Atmospheric Research, University of Hawai'i at Manoa, 1000 Pope Road, MSB 312, Honolulu, HI, 96822, USA*

<sup>2</sup>*NOAA/Pacific Islands Fisheries Science Center, 2570 Dole St., Honolulu, HI, 96822, USA*

## ABSTRACT

An approach to integrate sea surface temperature (SST) measurements into estimates of geolocations calculated by changes in ambient light level from data downloaded from pop-up satellite archival tags (PSAT) is presented. The model is an extension of an approach based on Kalman filter estimation in a state-space model. The approach uses longitude and latitude estimated from light, and SST. The extra information on SST is included in a consistent manner within the milieu of the Kalman filter. The technique was evaluated by attaching PSATs directly on thermistor-equipped global positioning system drifter buoys. SSTs measured in the PSATs and drifter buoy were statistically compared with SSTs determined from satellites. The method is applied to two tracks derived from PSAT-tagged blue sharks (*Prionace glauca*) in the central Pacific Ocean. The inclusion of SST in the model produced substantially more probable tracks with lower prediction variance than those estimated from light-level data alone.

**Key words:** blue shark, geolocation errors, Hawaii, Kalman filter, Lagrangian drifters, pop-up satellite archival tags, sea surface temperature

## INTRODUCTION

Fisheries research organizations, universities, and governments spend large amounts of funding on electronic tagging devices to address a variety of fundamental

biological and practical management questions (e.g. Arnold and Dewar, 2001; Gunn and Block, 2001). Electronic tagging devices are used to gain information about fisheries interactions, post-release survivability, and to delimit spawning locations and migration corridors. As water is (nearly) impenetrable to radio waves, alternative strategies are needed in electronic tagging devices to provide estimates of geographical location (i.e. geolocations) for fishes that do not regularly visit the surface. For animals that breath air, and hence are at the surface for long periods, it is possible to use Argos or Global Positioning System (GPS) tags, which are highly accurate. Archival or data storage tags and pop-up satellite archival tags (PSATs) are electronic tags that store ambient light-level data from which estimates of dawn and dusk can be used to calculate longitude (from local noon) and latitude (from local day length) (Wilson *et al.*, 1992; Hill, 1994; Musyl *et al.*, 2001).

Despite their enormous potential, archival tags and PSATs have several limitations, including tag retention, variable reporting rates, amount of transmitted data and the magnitude and extent of geolocation errors from light-based techniques. This paper provides a method that can be used to improve the accuracy of light-based geolocation estimates so that these may be applied to analyze movement in relation to oceanographic variability.

Raw geolocations (i.e. unfiltered and uncorrected) derived from light-based algorithms are often very noisy, and it is common that devices are mistakenly 'placed' hundreds of kilometers from their actual location (Gunn *et al.*, 1994; Musyl *et al.*, 2001; Sibert *et al.*, 2003). Estimation of latitude by light-based methods around the time of the equinox can be especially problematic because day length is nearly equal at all latitudes. Light-based geolocation methods are often rendered even more difficult by the behavior of the animals. For instance, animals such as swordfish and bigeye tuna spend a substantial fraction of the day at great depth where light intensity is near the threshold of sensitivity for most electronic tags, or dive near the time of sunrise and sunset (Musyl *et al.*, 2003). It is obvious that these raw geolocations are essentially useless by themselves because of the likely magnitude and extent of geolocation errors.

\*Correspondence. e-mail: anders.nielsen@hawaii.edu

Received 23 June 2004

Revised version accepted 18 May 2005

A widely used (e.g. Musyl *et al.*, 2003; Wilson *et al.*, 2005), and in most cases successful, approach to make sense of these raw geolocations is based on a state-space model (Sibert *et al.*, 2003), and estimated by the Kalman filter (Harvey, 1990). This model assumes that the tagged individual moves according to a biased random walk, and the raw geolocations are interpreted as the true position plus some measurement error. The measurement error is parameterized to produce large latitude errors near an equinox and smaller latitude errors near a solstice, which is often seen in the raw geolocations as a side effect of light-based geolocation (Hill and Braun, 2001). State-space models are also used in animal tracking in combination with the Bayesian paradigm, where prior beliefs are part of the specified model (Jonsen *et al.*, 2003).

The Kalman filter is named after Rudolph Emil Kalman, who published this elegant recursive method (Kalman, 1960). The Kalman filter is one of the most well-known and often-used tools in statistical time-series analysis, with a wealth of applications in, for instance, finance, missile guidance, and interactive computer graphics. The usages of the Kalman filter are described in full detail in the Methods section, but intuitively this approach uses two components: raw geolocations and predictions from the underlying random walk. The method uses the raw geolocations to estimate the random walk parameters in the periods when data are reliable, but relies mainly on the random walk predictions in periods where data are unreliable.

However, most electronic tagging devices store more than just light intensities. Water temperature and pressure (depth) are also commonly stored. Ancillary information can be used to estimate geographic position. In particular, Smith and Goodman (1986) envisaged the use of ambient water temperature to estimate latitude. The use of sea surface temperature (SST) to estimate latitude from archival tags has since become common (e.g. Inagake *et al.*, 2001; Takahashi *et al.*, 2003; Teo *et al.*, 2004).

The data stored in electronic tags represent a substantial investment of time and financial resources. Temperature measurements are already stored in the tags, and using them to refine the geolocation estimates from the tags is a logical extension. The purpose of this paper is to extend a model relying only on raw geolocations to a model that uses all the information available from the tags – longitude, latitude and SST. The method developed in this paper is not restricted to SST, but could potentially be used to reconstruct tracks from other measurements (e.g. depth and temperature at depth) if they were available.

## MATERIALS AND METHODS

Wildlife Computers pop-up archival transmitting tags (PAT, version 2) were deployed on two WOCE/GPS drifter buoys (ClearSat-15 from Clearwater Instrumentation, Inc., Watertown, MA, USA), and set to pop up after 9 months. The tags were attached with short monofilament tethers (approximately 40 cm) so that they would float at the surface without being shaded by the bouys for extended periods. Both drifter buoys were deployed (PAT 21760 at 161.975 W, 23.086 N and PAT 21765 at 174.538 W, 26.304 N) in September 2002 prior to the equinox and set to acquire hourly histograms of depth and temperature.

As part of ongoing investigations to determine post-release survivability in blue shark (*Prionace glauca*) released from longline fishing gear, model PTT-100 PSATs from Microwave Telemetry (MT) were affixed to two blue sharks in the central Pacific Ocean (M.K. Musyl and R.W. Brill, unpublished data). In these studies, sharks were caught by conventional longline fishing gear configured in typical Hawaiian 'swordfish' style (i.e. shallow nighttime sets with less than approximately 100 m, as determined by attached time depth recorders, about 4–5 hooks per basket with a green chemical light stick above each dropper, usually baited with squid (*Illex* spp.) on 15/0 hooks. Sharks were brought on board the NOAA research vessel *Townsend Cromwell* in a sling and were restrained by crew with mattresses. PSATs were affixed to the shark's dorsal fin by drilling a hole (approximately 10–15 mm diameter) near the base of the fin and threading 49-braid stainless steel wire encased in Tygon tubing which acted as the harness. Stainless crimps were used to attach the harness to a tether made of 122 kg fluorocarbon which was crimped to the nose cone of the PSAT.

Pop-up satellite archival tags from MT deployed on sharks were programmed to acquire raw temperature and pressure (depth) readings every hour and the tag's pop-off dates were set at 13 months after deployment. 'Fail-safe' options were programmed in the PSAT to receive data in the event the animal died and sank or if the tag was shed. In the event of a sinking PSAT, a pressure release mechanism (corrosional link) was set at 1200 m. In this scenario, the tag jettisons from the shark and floats to the surface to start transmitting data to Argos satellites. Alternatively, if the tag does not experience any significant pressure changes in four consecutive days (e.g. shed tag floating at the surface or a stationary tag at a depth <1200 m), it will automatically initiate data recovery procedures. For the MT tags, estimates for dawn and dusk are

automatically calculated in the tag by a proprietary algorithm (Gunn and Block, 2001). PSAT 13097 was affixed to an approximately 2-m female blue shark on April 2, 2001 at 158.76 W, 28.28 N, and PSAT 13093 was affixed to a 1.5-m female blue shark on April 11, 2001 at 158.28 W, 18.88 N.

Two satellite-derived SST products were obtained for inclusion in the Kalman filter. Fine-scale SST fields (error estimate 0.3–0.5°C) at a resolution of 9 km and 8-day averages were obtained from Advanced Very High Resolution Radiometers (AVHRR) Pathfinder data (Vazquez et al., 1998). Mesoscale SST fields (error estimate 0.7°C) at geographic resolution of 1° and a weekly time scale were based on an optimal interpolation analysis of AVHRR and *in situ* ship and buoy data (Reynolds and Smith, 1994). Fine-scale data were used for the blue shark tags. Mesoscale data were used for the drifter buoy tags as these data have no gaps caused by clouds. The rationale for using the meso-scale product was to simplify the computations, because the geolocation estimates for the buoys were so poor.

*Model*

The model is based on an improved and extended version of the state-space Kalman filter model of Sibert et al. (2003), where the state equation is describing the movements along the sphere. A random walk model is assumed:

$$\alpha_i = \alpha_{i-1} + c_i + \eta_i, \quad i = 1, \dots, T. \quad (1)$$

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

The measurement equation of the state-space model is a non-linear function describing the expected observation at a given state  $(\alpha_i)$ . Each observation  $y_i$  consists of three elements: longitude, latitude, and SST. The first two coordinates are the light-based geolocation estimates and the last is the SST recorded by the tag. The measurement equation describing  $y_i$  is:

$$y_i = z(\alpha_i) + d_i + \varepsilon_i, \quad i = 1, \dots, T. \quad (2)$$

The first two coordinates of  $z$  comprise the coordinate change function, and the last coordinate describes the expected SST at a given position.  $z$  is given by:

$$z(\alpha_i) = \begin{pmatrix} \frac{\alpha_{i,1}}{60 \cos(\alpha_{i,2}\pi/180/60)} \\ \frac{\alpha_{i,2}}{60} \\ \tau\left(\frac{\alpha_{i,1}}{60 \cos(\alpha_{i,2}\pi/180/60)}, \frac{\alpha_{i,2}}{60}\right) \end{pmatrix}. \quad (3)$$

Here the factor  $\pi/180$  converts from degrees to radians, and 60 is the distance corresponding to 1° of longitude at the equator. The function  $\tau(\text{longitude, latitude})$  describes the expected SST at a given location and time. How this function is constructed will be described later. The observational bias  $d_i = (b_{lon}, b_{lat}, b_{SST})'$ , if present, describes systematic measurement errors, for instance if the internal clock in the tag is not absolutely correct. The bias term was not necessary in any of the tracks presented in this paper. The measurement error  $\varepsilon_i$  is assumed to follow a Gaussian distribution with mean vector 0 and covariance matrix  $H_i$ , where

$$H_i = \begin{pmatrix} \sigma_{lon}^2 & 0 & 0 \\ 0 & \sigma_{lat_i}^2 & 0 \\ 0 & 0 & \sigma_{SST}^2 \end{pmatrix}. \quad (4)$$

The variance of the latitude measurements  $\sigma_{lat_i}^2$  is closely related to the equinox, as measurements close to an equinox have large latitude errors. The following variance structure is assumed:

$$\sigma_{lat_i}^2 = \frac{\sigma_{lat_0}^2}{\cos^2\{2\pi[J_i + (-1)^{s_i}b_0]/365.25\} + a_0}, \quad (5)$$

where  $J_i$  is the number of days since last solstice prior to all observations,  $s_i$  is the season number since the beginning of the track (1 as long as  $J_i < 182.625$ , then 2 for the next 182.625 days, then 3 and so on). The model parameter  $b_0$  express the number of days, before or after alternating with season, the latitude errors will peak compared with the date of the equinox. The model parameter  $\sigma_{lat_0}^2$  is the minimum latitude variance, and  $a_0$  ensures an upper bound for the latitude variance.

*Extended Kalman filter*

As the model described in the previous section is non-linear, an approximation is needed to apply the Kalman filter. The extended Kalman filter (Harvey, 1990) simply uses a first-order Taylor approximation around the optimal estimator  $\hat{\alpha}_{i|i-1}$ :

$$z(\alpha_i) \approx z(\hat{\alpha}_{i|i-1}) + \hat{Z}_i(\alpha_i - \hat{\alpha}_{i|i-1}). \quad (6)$$

Here  $\hat{Z}_i$  is the first derivative (or the Jacobi matrix) of the function  $z$ , which is calculated as:

$$z'(x_i) = \frac{\partial z(x_i)}{\partial x_i} = \begin{pmatrix} \frac{1}{60 \cos(x_{i,2} \cdot \pi / 180 / 60)} & \frac{x_{i,1} \pi \sin(x_{i,2} \cdot \pi / 180 / 60)}{180 [60 \cos(x_{i,2} \cdot \pi / 180 / 60)]^2} \\ 0 & \frac{1}{60} \\ \frac{\partial \tau}{\partial x_{i,1}} \frac{1}{60 \cos(x_{i,2} \cdot \pi / 180 / 60)} & \frac{\partial \tau}{\partial x_{i,1}} \frac{x_{i,1} \pi \sin(x_{i,2} \cdot \pi / 180 / 60)}{180 [60 \cos(x_{i,2} \cdot \pi / 180 / 60)]^2} + \frac{\partial \tau}{\partial x_{i,2}} \frac{1}{60} \end{pmatrix} \quad (7)$$

The optimal estimator  $\hat{\alpha}_{i|i-1}$  is inserted into the matrix in Eqn 7 to obtain  $\hat{Z}_i$ . The extended Kalman filter update equations are written as:

$$\hat{\alpha}_{i|i-1} = \hat{\alpha}_{i-1} + c_i, \quad \text{predict next position,} \quad (8)$$

$$P_{i|i-1} = P_{i-1} + Q_i, \quad \text{calculate its variance,} \quad (9)$$

$$F_i = \hat{Z}_i P_{i|i-1} \hat{Z}_i' + H_i, \quad \text{transform into measurement variance,} \quad (10)$$

$$w_i = y_i - z(\hat{\alpha}_{i|i-1}) - d_i, \quad \text{calculate prediction error,} \quad (11)$$

$$\hat{\alpha}_i = \hat{\alpha}_{i|i-1} + P_{i|i-1} \hat{Z}_i' F_i^{-1} w_i, \quad \text{adjust prediction with observation,} \quad (12)$$

$$P_i = P_{i|i-1} - P_{i|i-1} \hat{Z}_i' F_i^{-1} \hat{Z}_i P_{i|i-1}, \quad \text{adjust prediction variance.} \quad (13)$$

The filter is started by calculating  $\hat{\alpha}_0 = z^{-1}(y_0)$  and assuming this position to be known without error ( $P_0 = 0_{2 \times 2}$ ).

Model parameters are estimated by maximum likelihood; the negative log-likelihood function for the Kalman filter is (Harvey, 1990):

$$\ell(\theta) = -\log L(Y|\theta) = \frac{3T}{2} \log(2\pi) + 0.5 \sum_{i=1}^T \log(|F_i|) + 0.5 \sum_{i=1}^T w_i' F_i^{-1} w_i. \quad (14)$$

*Obtaining the SST field  $\tau$  and its derivative*

The temperature field  $\tau$  and its derivatives  $\partial \tau / \partial x_{i,1}$  and  $\partial \tau / \partial x_{i,2}$  must be known in order to evaluate the coordinate change function (Eqn 3). Naturally the temperature field is not known, but must be estimated

from observations. A fine-scale SST field is comprised of approximately 200 000 measurements. Such a substantial volume of data could probably be handled directly in the Kalman filter numerical optimization, but would require a substantial amount of computer memory or clever data structures.

Smoothing of observed SST field is appropriate for this work for several reasons. The Kalman filter requires that  $\tau$  and its derivatives be evaluated at all points within the study area. Observations are often missing in some areas because of cloud cover. The observed SST field may be very irregular so that the apparent SST may vary substantially over short distances because of fine structure in the temperature field or to measurement noise. Finally, the daily SST estimate computed from the data recorded by the tag is in fact a daily average, so even if the true SST field is not smooth, the tag measurement corresponds to a field which has been locally smoothed by averaging.

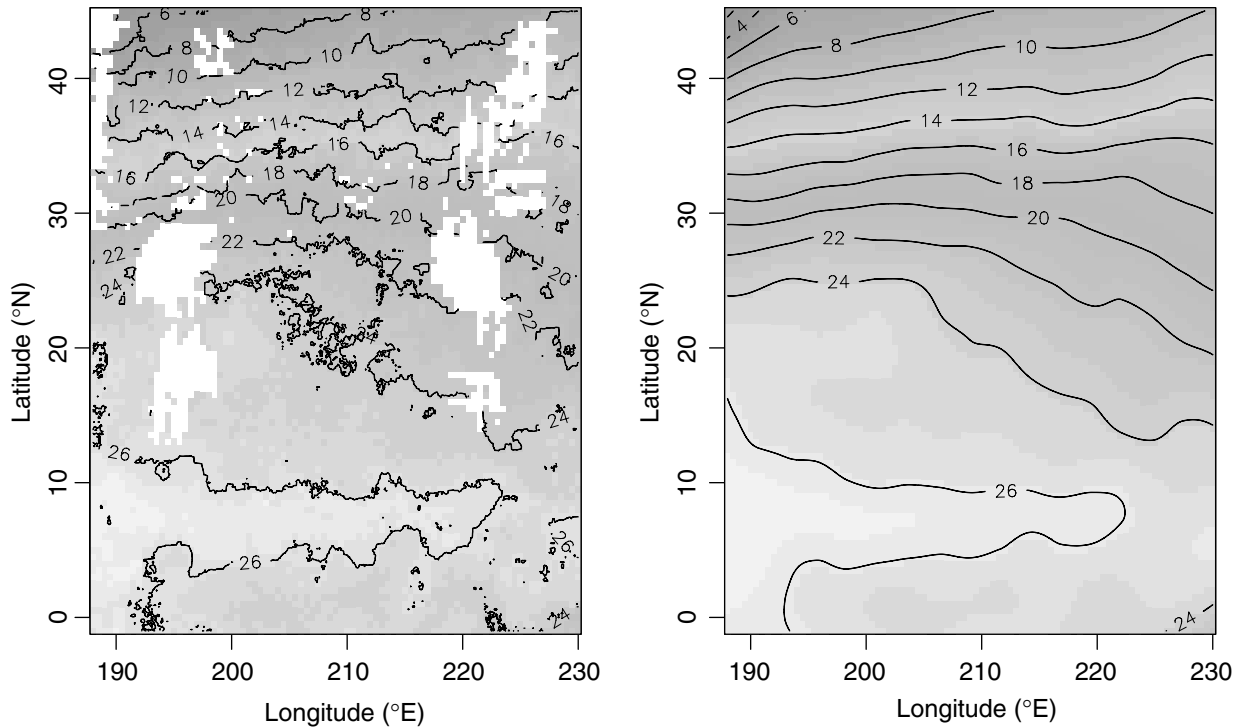
For the purposes of the Kalman filter, the estimated SST field is represented by a local polynomial regression (Loader, 1999). This algorithm fits local polynomials to the observations to ensure a surface smoothed to a degree specified by the modeler. The ‘nearest neighbor fraction’ determines what fraction of the nearest points are used to fit the local polynomial. For the tracks analyzed here, a nearest neighbor fraction of 5% was chosen. Three different degrees of smoothing were tested (2.5%, 5%, 10%) to investigate the sensitivity to the degree of smoothing. Advantages of this algorithm are that data can be smoothed to the extent necessary, the derivative field is computed directly by the smoothing algorithm, and predictions can be made where data are missing. Figure 1 shows an example of a smoothed SST field.

*The most-probable track*

The 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:  $\hat{\alpha}_i = E(\alpha_i | y_1, \dots, y_i)$ . Better estimates can be produced once the entire track is known. The most-probable track is calculated using all observations after all parameters have been estimated:  $\hat{\alpha}_{i|T} = E(\alpha_i | y_1, \dots, y_T)$ .

The actual computation of the most-probable track is made in a single backwards updating sweep of the predicted track. 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 known  $\hat{\alpha}_{T|T} = \hat{\alpha}_T$  and the

**Figure 1.** The observed fine-scale sea surface temperature field from the first week of January 2001 (left), and the (5% nearest neighbor fraction) smoothed version used by Kalman filter (right).



prediction error of the last point  $P_{T|T} = P_T$ . The following equations are used to compute the previous points of the most-probable track:

$$\hat{\alpha}_{i|T} = \hat{\alpha}_i + P_i^*(\hat{\alpha}_{i+1|T} - \hat{\alpha}_i - c_{i+1}),$$

update position given all observations, (15)

$$P_{i|T} = P_i + P_i^*(P_{i+1|T} - P_{i+1|i})P_i^{\prime}, \text{ where } P_i^* = P_i P_{i+1|i}^{-1}.$$

(16)

This technique is known as ‘smoothing’ in textbooks on the Kalman filter (e.g. Harvey, 1990), as the resulting track most often is smoother than the predicted track. The matrix  $P_{i|T}$  is the covariance estimate of the  $i$ th position estimate on the most-probable track measured in nautical miles.  $P_{i|T}$  is a measure of the variability of the position estimate after all observations (past and future) have been taken into account.

*Fixed pop-up position*

If the last position is known without error, or with an error term which is negligible compared with the light-based geolocations, it can be used as a fixed point. This corresponds to setting the measurement error and the measurement bias to zero for the last observation

(( $H_T$ )<sub>1,1</sub>, ( $H_T$ )<sub>2,2</sub>, ( $d_T$ )<sub>1</sub>, and ( $d_T$ )<sub>2</sub> all zero). This results in a predicted (and most-probable) track ending up in the fixed pop-up location.

A software package to run the model described has been developed and is publicly available (Nielsen and Sibert 2005).

**RESULTS**

The tags on both drifter tags reported on schedule. PAT 21760 should have transmitted 6024 readings for both depth and temperature. For the raw depth histogram data, only 107 readings were received (1.8% of all possible readings). For the raw temperature histogram data, only 96 readings were received (1.6% of all possible readings). For the profile of depth-temperature data (PDT), only 92 readings were received (1.5% of all possible readings). From the PDT data the average of the daily SST temperature readings at the surface were taken. Software from the manufacturer (Wildlife Computers, Redmond, WA, USA; dated March 2003) was used to estimate geolocations for the PSAT and 167 estimates were given (66.5% of all possible days). As a comparison, the GPS drifter buoy over the same time period transmitted 7316 geolocations (more than one per hour) and 1166 SSTs (or 19.4%).

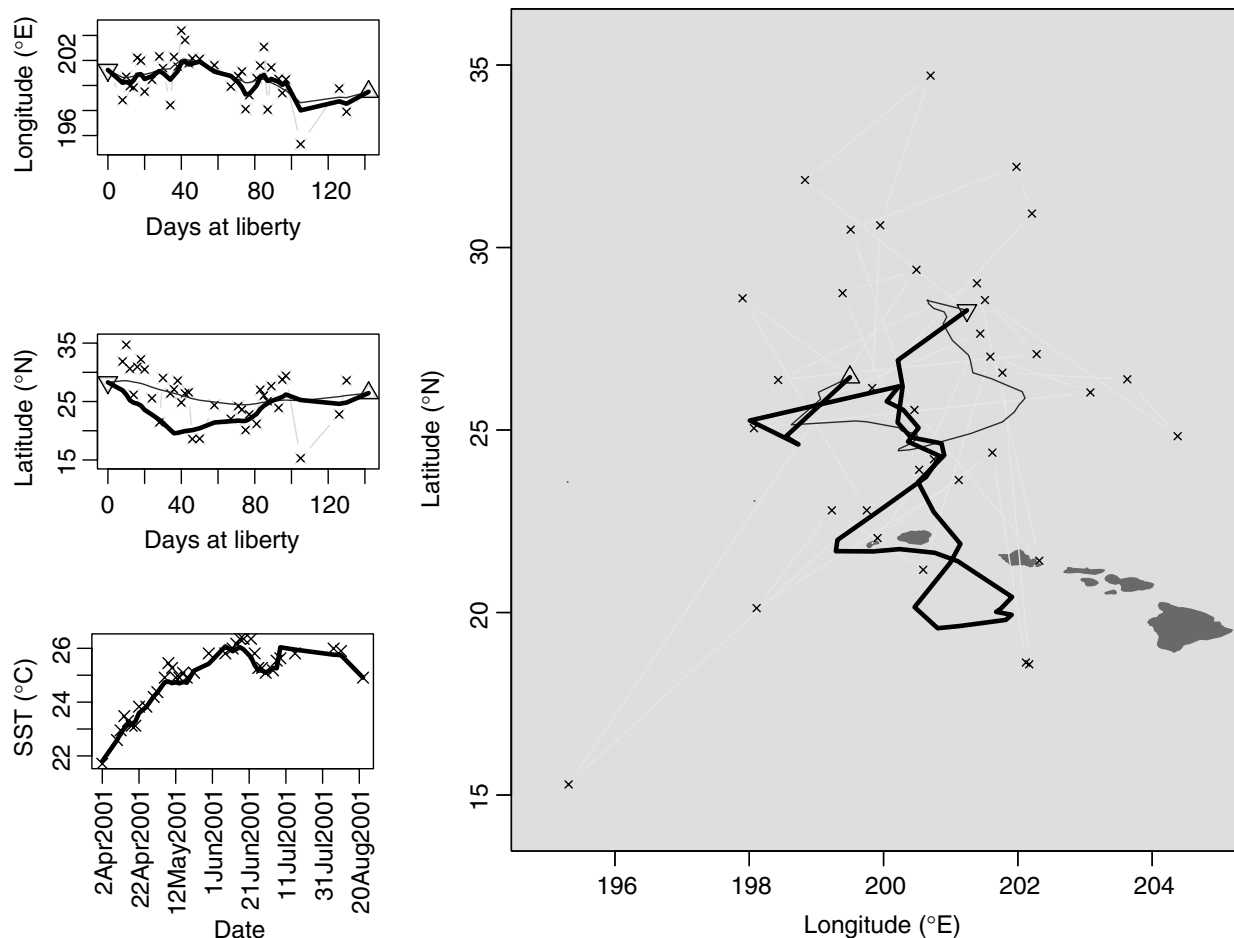
PAT 21765 should have transmitted 5856 readings for both depth and temperature. For the raw depth histogram data, only 81 readings were received (1.4% of all possible readings). For the raw temperature histogram data, only 71 readings were received (1.2% of all possible readings). For the PDT data, only 92 readings were received (1.6% of all possible readings). Software from the manufacturer was used to estimate geolocations for the PSAT and 154 estimates were given (63% of all possible days). As a comparison, the GPS drifter buoy over the same time period transmitted 7521 geolocations (more than one per hour) and 1234 SSTs (or 21.1%).

The shark affixed with PSAT 13093 spent 90% of its depth at 100 m or less. The shark with tag 13097 spent 90% of its time at 160 m or less and had a high correlation between average nighttime depth and

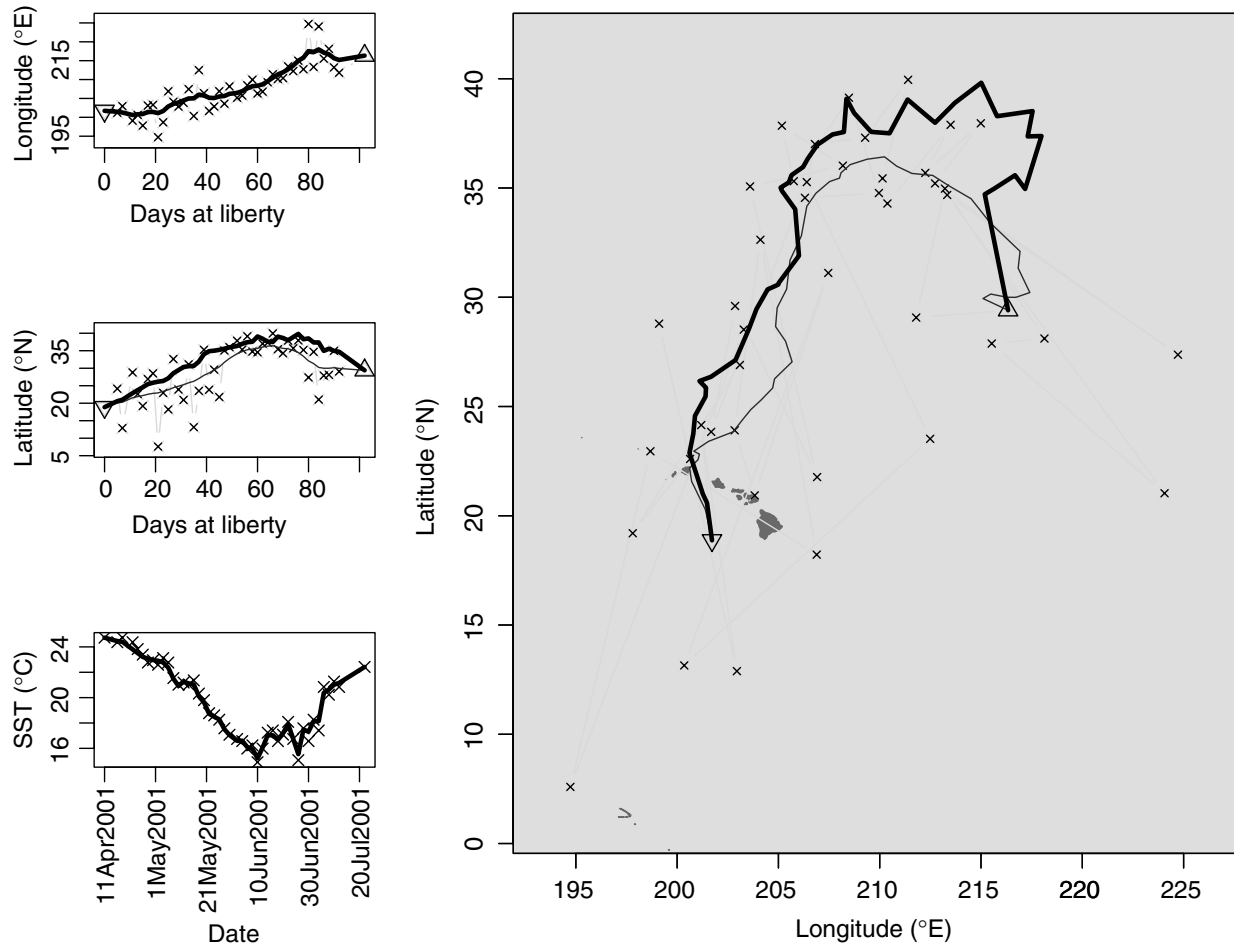
lunar illumination indicating a pronounced diel depth pattern. For PSAT 13097, at liberty for 233 days, 47% of the possible raw depth and temperature data and 42% of geolocations were received. For PSAT 13097, at liberty for 102 days, 50% of the possible raw depth and temperature data and 42% of geolocations were received.

Analysis of the two-tagged female blue sharks with the method using only the light-based geolocations and the extended method with SST showed pronounced differences in the estimated, most-probable tracks (Figs 2 and 3). The longitude estimates for the most-probable tracks remained fairly unchanged by the inclusion of SST. The latitude estimates of the most-probable tracks were dramatically changed by inclusion of SST (Figs 2 and 3). These results are not surprising as longitudinal SST gradients are small

**Figure 2.** Most-probable track for blue shark number 13097 (right) fitted with (thick line) and without sea surface temperature (SST) information (thin line). Thin white line connects the raw light-based geolocations. Left column shows how well the model fits the three different information sources (longitude, latitude, and SST). Raw geolocations and observed SST are marked by crosses, deployment point is marked by '▽', and known recapture/pop-up position is marked by '△'.



**Figure 3.** Most-probable track for blue shark number 13093 (right) fitted with (thick line) and without sea surface temperature (SST) information (thin line). Thin white line connects the raw light-based geolocations. Left column shows how well the model fits the three different information sources (longitude, latitude, and SST). Raw geolocations and observed SST are marked by crosses, deployment point is marked by '∇', and known pop-up position is marked by '△'.



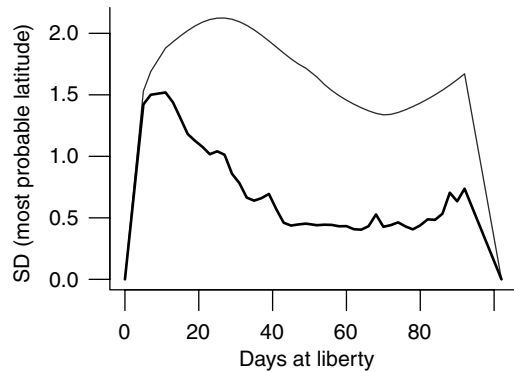
while latitudinal SST gradients are large in this region. The variation of the latitude observations (second plot in left column) around the most-probable track is much larger than the variation of the tag's SST measurements compared with the remotely sensed SSTs along the most-probable track (third plot in left column).

To quantify the difference between tracks estimated with and without SST measurements, the covariance of each point on the most-probable track was calculated from the covariance matrix  $P_{i|T}$  in Eqn 16. The resulting covariance for each point along the track matrix is transformed into degrees of longitude and latitude by the delta method,  $V_i = \hat{Z}_i P_{i|T} \hat{Z}_i'$ .  $V_i$  is now an estimate of the covariance of the  $i$ th point on the most-probable track expressed in degrees of longitude and latitude. This variance could for instance

be used to construct confidence intervals (ellipses) for each point on the track.

The longitude variance  $(V_i)_{1,1}$  of both tracks was not changed much by including SST. This was expected, as the SST field is fairly constant longitudinally. For track number 13097 the latitude variance  $(V_i)_{2,2}$  was also very similar. The average latitude variance with SST was 0.73 compared with 0.75 without SST. This was slightly surprising. A closer look revealed that this blue shark was swimming at latitudes where the temperature gradient was not very steep (compare Fig 2 with 1). For track number 13093, the latitude variances were much lower when SST measurements were included. The average latitude variance with SST was 0.59 compared with 2.9 without SST. The latitude standard deviation for each point on the most-probable track estimated with and

**Figure 4.** Estimated standard deviation for the latitude component of the most-probable track 13093 estimated with (lower thick line) and without sea surface temperature information (upper thin line).



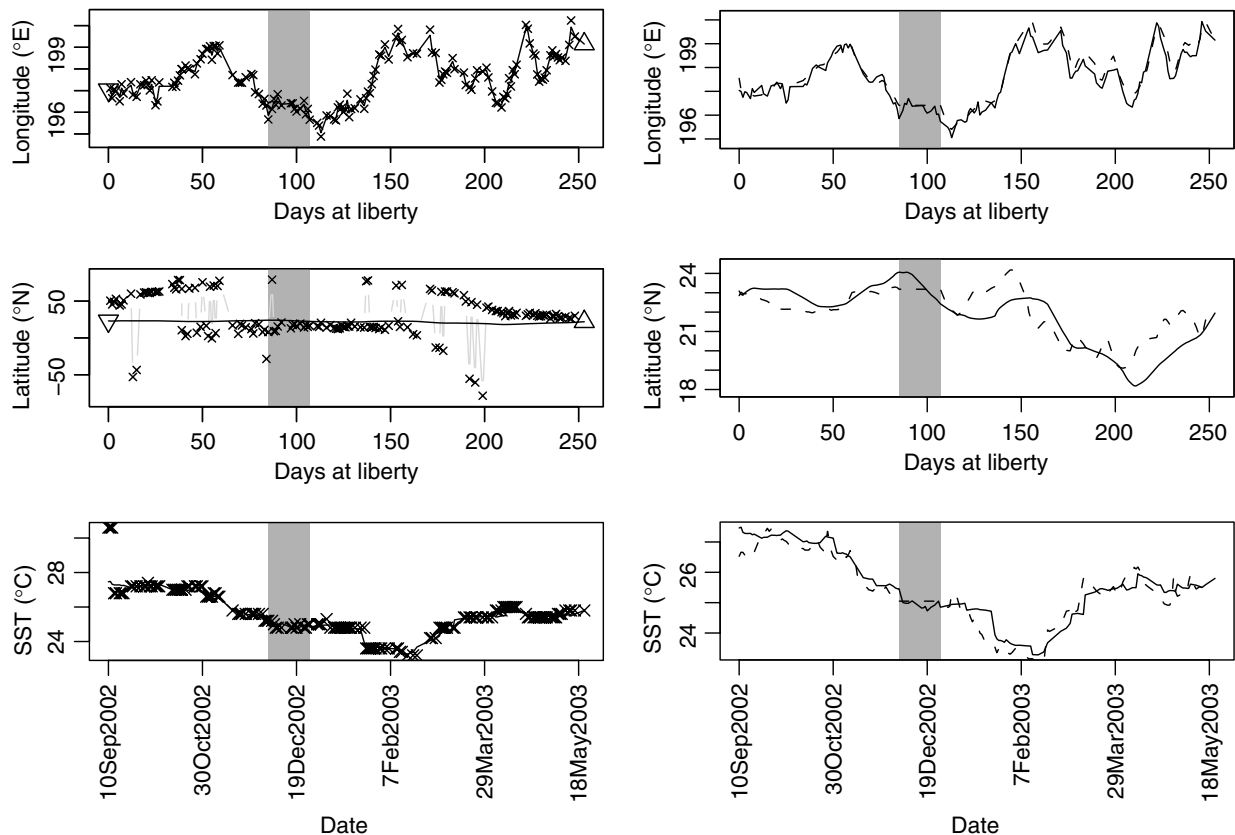
without SST measurements (Fig. 4) illustrates the difference. The latitude standard error of the purely light-based method is between 1.5° and 2°, meaning that a 95% confidence interval could span up to 8°.

The standard error of the method including SST is between 0.4 and 1.5, and about half of the time around 0.5, which corresponds to a 95% confidence interval spanning only 2°. The average longitude variance was 2.3 for track number 13093 and 0.54 for track number 13097, both estimated in the model with SST.

The double-tagging buoy experiment was not designed to evaluate the method of incorporating SST, but to investigate the error structure of light-based geolocations. Unfortunately the light-based latitude information from these specific tags produced estimates with huge geolocation errors (Figs 5 and 6). With measurement errors of up to 50° latitude, this information is practically useless. However, the double-tagging buoy experiment offers a unique opportunity to test the SST method in the absence of consistent latitude information on a moving tag.

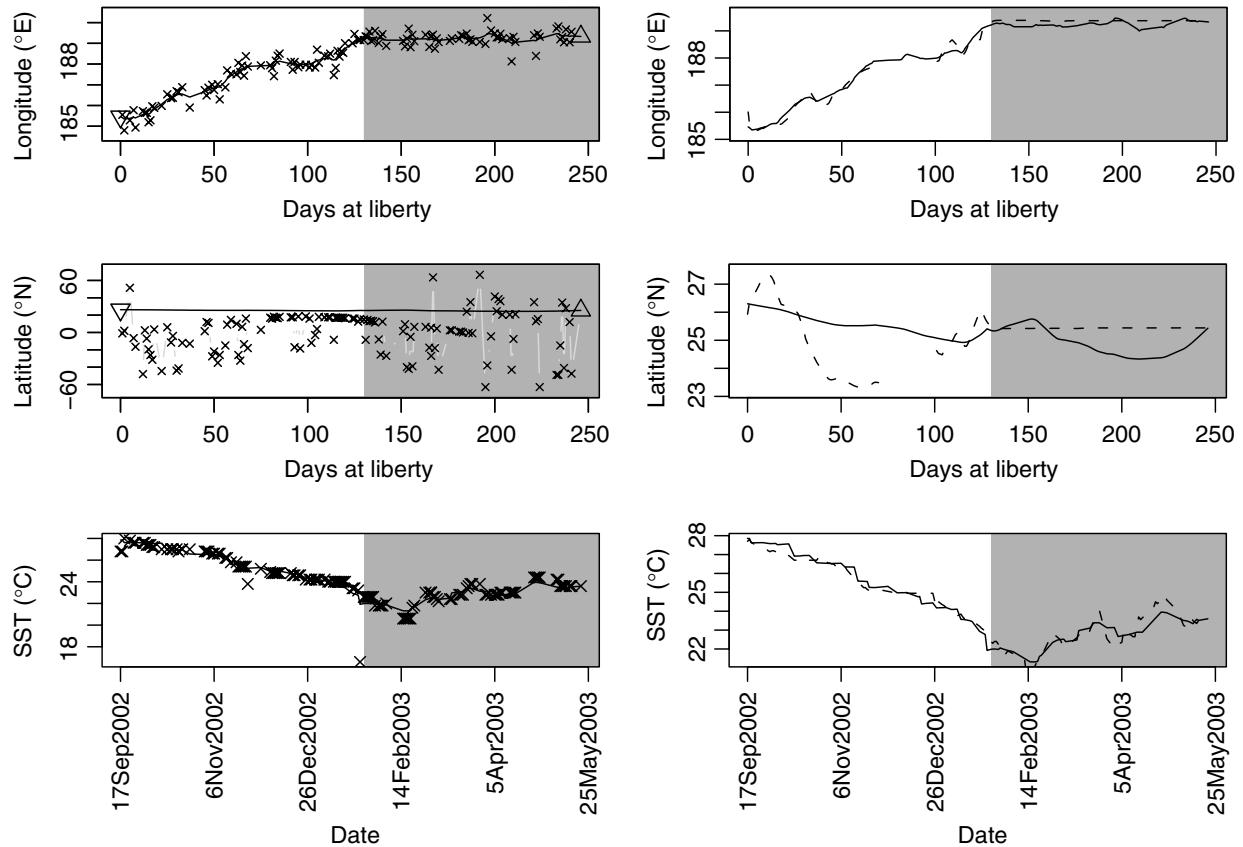
The first double-tagging buoy experiment (Fig. 5) showed that the SST-enhanced method is able to estimate the actual track of the drifter buoy. The estimated longitude coordinates from the most-probable track match the GPS-determined longitude very

**Figure 5.** Most-probable track for drifter buoy 21760 (solid line) compared with the raw observations (crosses left plot), and to the true positions from global positioning system measurements (dashed lines right plot). The shaded area indicates a time period when the buoy was beached on an island.





**Figure 6.** Most-probable track for drifter buoy 21765 (solid line) compared with the raw observations (crosses left plot), and to the true positions from GPS measurements (dashed lines right plot). The shaded area indicates a time period when the buoy was caught on a reef.



closely, as expected (the root mean squared, or RMS, difference =  $0.31^\circ$ ) and are slightly better than the raw longitude (RMS difference =  $0.36^\circ$ ). The latitude coordinates are also close to the GPS-determined latitude (RMS difference =  $0.99^\circ$ ), which is more remarkable, as the raw latitude was not very helpful (RMS =  $29.23^\circ$ ). The most-probable latitude track is derived almost solely from SST information. The buoy beached itself for a period of 3 weeks and then went back into the water (Fig. 5). The buoy kept transmitting and recording, and the reconstructed track seems unaffected by stranding.

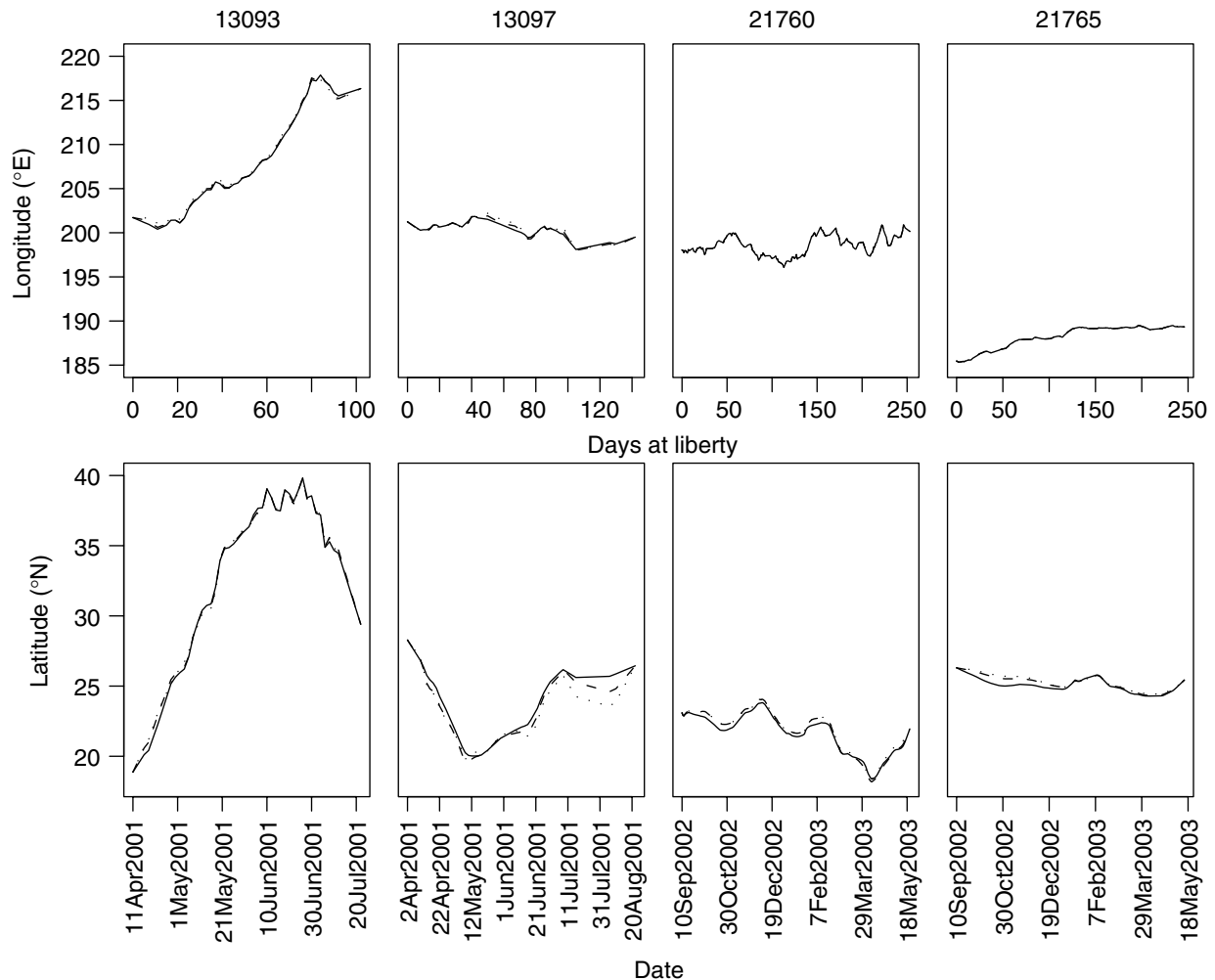
The second double-tagging buoy experiment (Fig. 6) is more problematic to evaluate, as the buoy was caught on a reef about half of the time. The most-probable longitude track agrees well with the GPS information (RMS difference =  $0.15$ ) and is better than the raw longitude estimates (RMS difference =  $0.32$ ). Raw latitude measurements had errors up to  $60^\circ$  (RMS difference =  $36.17$ ) compared with the actual track. The most-probable latitude track is off by only

$1^\circ$  or  $2^\circ$  from the actual track and had smaller RMS differences ( $1.00$ ). The buoy entered Maro Reef after 120 days at liberty and remained within the reef complex for the last 4 months (Fig. 6). The PSAT functioned normally during this period recording data and transmitting on schedule.

To illustrate the sensitivity of the degree of smoothing to the SST temperature field, three different degrees of smoothing were tested on the four tracks. Normal (5% nearest neighbor fraction), twice as coarse (10%), and twice as fine (2.5%). Here fine refers to less smoothing, which is closer to the observed SST field. This sensitivity analysis showed only minor differences between the reconstructed tracks (Fig. 7). In fact, most of the tracks were identical. In one case, the very finest smoothing seemed to give a more variable track in a few points.

The degree of smoothing used in the track analysis was chosen simply by looking at the smoothed fields compared with the observed fields. A more objective

**Figure 7.** Most-probable longitude (top row) and latitude (bottom row) of the four estimated tracks with three levels of smoothing of the sea surface temperature (SST) field: normal (dashed line), twice as coarse (solid line), and twice as fine (dotted line). Fine-scale SST fields were applied to tracks 13093 and 13097. Mesoscale data were applied to tracks 21760 and 21765.



and automatic approach would be to select the degree of smoothing based on a cross-validation scheme (Efron and Tibshirani, 1993).

## DISCUSSION

Geolocation based on indirect measurements is unlikely to ever achieve the precision of GPS methods, but combining several indirect measurements in a coherent manner is the best solution possible. The latitude variance estimates,  $(V_i)_{2,2}$ , are substantially lower than those anticipated by Smith and Goodman (1986) and approach the theoretical limit discussed by Metcalfe (2001).

The general approach of matching individual measurements to a corresponding externally determined field, within the milieu of the Kalman filter, can

potentially be used to further improve (or replace) light-based geolocation by including more measurable or detectable environmental variables. For instance salinity, geomagnetism, chemical tracers, or water depth for demersal species (Hunter *et al.*, 2003). The present paper demonstrates the use of supplementary fields by illustrating how the inclusion of SST measurements is able to reduce the geolocation error substantially (Fig. 4), and even reconstruct a track where the light-based latitude measurements were practically useless (Fig. 5). Reduction in geolocation error is greatest in the directions and areas where the gradient of the ancillary field is steepest. In the case of SST, the SST field in the study area had steepest gradients in the latitude direction, which is also the direction where light-based geolocation is poorest. When considering to include additional variables to improve

geolocation, it should be a concern whether or not that information is really already implicitly represented. For instance, if the SST and salinity fields are proportional in the study area, it would not contribute much to include them both. The state space Kalman filter approach (Sibert *et al.*, 2003) is an important part of the presented method. It ensures that the matching between tag observations and the underlying field is made locally. The field can potentially have many areas where the value matches the tag observed value, and the Kalman filter provides an objective method to select the value that is most consistent with all observations.

Drifter buoys are not the perfect tools for evaluating how well this method works on live animals, as their behavior is somewhat different. Free-ranging large pelagic fish make vertical excursions in the water column and do not run aground. These abnormal periods could have influenced the SST measurements. The 3-week beaching seems not to have influenced the measurements or the reconstructed track. The 4 months at a fixed position in the shallow waters of a reef might have given slightly higher SST readings than if a tagged individual was swimming freely in the surrounding waters. This may account for the more southerly estimated track. Despite these difficulties the drifter buoy experiments demonstrate that SST is useful in estimating latitude, even when the light-based latitude is practically useless.

The inclusion of ocean temperature in the Kalman filter estimation process could result in more accurate geolocations for several pelagic species. Light-based geolocations are problematic for species with large diurnal vertical migrations (e.g. swordfish, bigeye tuna and thresher sharks). These species descend before dawn to occupy mesopelagic depths during the day and ascend after dusk to occupy near-surface depths at night. Latitude estimates are often inaccurate because of dim ambient light at mesopelagic depths and extensive vertical movements at the time of dawn and dusk (Musyl *et al.*, 2001). However, these species often occupy the mixed layer during the night and SST information could improve the geolocations. Takahashi *et al.* (2003) used a searching algorithm to match tag temperatures at 0, 80 and 160 m to ocean temperatures at a monthly 2° latitude and 5° longitude scale to estimate latitudes from an archival tag recovered from a swordfish. A similar approach could be envisaged in a Kalman filter process whereby tag temperatures at several depths could be linked with observed temperature-at-depth data or Oceanic General Circulation Model (OGCM) model output.

A simple sensitivity analysis was conducted to illustrate the effect of smoothing the SST temperature field. While current results indicate only minor differences when using three different degrees of smoothing, further work is required to investigate the spatial as well as temporal effects of scale and smoothing. The appropriate scale of SST data and smoothing will depend on several factors such as the ocean dynamics in the study area and data quality because of cloud cover. Scale and smoothing considerations would be different for an analysis of a western boundary current (Kuroshio, Gulf Stream) region compared with a less stratified mid-ocean gyre. The abundance of clouds may dictate the scale and smoothing required as the Kalman filter requires that all points be included in the study area and smoothing was necessary to remove SST pixels with no data (excessive cloud cover), especially in the fine-scale (9 km) data.

Use of the average of the water temperature measurements collected from the tag at depths <15 m is an arbitrary choice. Using for instance the median or the mode might work equally well (or better) within the framework of the outlined method. Similarly, other sources and scales of SST fields should be considered.

The application of SST to estimation of positions from archival tag data inevitably requires some arbitrary decisions: a specific SST product, scale, and resolution must be selected; a means to estimate SST from data recorded by the tag must be devised; a criterion to determine what constitutes a match between environmental and tag SST estimates must be specified. Modifications to the Kalman filter model described here eliminate the third decision by including the SST variance in the likelihood function (Eqn 14) in a statistically consistent way with other sources of variance. Furthermore, smoothing the SST surface and the use of averaging to compute the tag SST mitigate the consequences of the first two decisions. Other temperature correction methods assume that the light-based longitude estimates are either error free or have errors that can be reduced to near zero by removal of outliers. The Kalman filter makes no such assumptions and uses all of the data. In principal, inclusion of SST also enables correction of longitude errors where tracks cross regions with large longitudinal SST gradients.

#### ACKNOWLEDGEMENTS

The research was supported by the SLIP research school under the Danish Network for Fisheries and Aquaculture Research (<http://www.fishnet.dk>) financed by the Danish Ministry for Food, Agriculture and Fisheries and the Danish Agricultural and

Veterinary Research Council. The work described in this paper was partly sponsored by the University of Hawaii Pelagic Fisheries Research Program under cooperative agreement number NA17RJ1230 from the National Oceanic and Atmospheric Administration. We thank Kevin Wong, NOAA/PIFSC and Phil White, crew, and officers of the NOAA research ship *Townsend Cromwell* for their outstanding help in deploying the drifter buoys and PSAT tags. We thank the NOAA CoastWatch – Central Pacific Node for providing SST data. Finally we acknowledge Steven Teo for comments and suggestions to a previous version of this paper, which greatly improved the final result.

## 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*. J.R. Sibert & J.L. Nielsen (eds) Dordrecht: Kluwer Academic Press, pp. 7–64.
- Efron, B. and Tibshirani, R.J. (1993) *An Introduction to the Bootstrap*. London: Chapman & Hall.
- Gunn, J. and Block, B. (2001) Advances in acoustic, archival, and satellite tagging of tunas. In: *Tuna Physiology, Ecology, and Evolution*. B.A. Block & E.D. Stevens (eds) New York: Academic Press, pp. 167–224.
- Gunn, J., Polachek, T., Davis, T., Sherlock, M. and Betlehem, A. (1994) The development and use of archival tags for studying the migration, behaviour and physiology of bluefin tuna, with an assessment of the potential for transfer of the technology for groundfish research. *ICES C.M. Mini* 21:1–23.
- Harvey, A.C. (1990) *Forecasting, Structural Time Series Models and the Kalman Filter*. Cambridge: Cambridge University Press.
- Hill, R. (1994) Theory of geolocation by light levels. In: *Elephant Seals: Population Ecology, Behaviour, and Physiology*. B.J., LeBoeuf & R.M., Laws (eds) Berkeley, CA: University of California Press, pp. 227–236.
- Hill, R.D. and Braun, M.J. (2001) Geolocation by light level – the next step: latitude. In: *Electronic Tagging and Tracking in Marine Fisheries*. J.R. Sibert & J. Nielsen (eds) Dordrecht: Kluwer Academic Publishers, pp. 315–330.
- Hunter, E., Aldridge, J.N., Metcalfe, J.D. and Arnold, G.P. (2003) Geolocation of free-ranging fish on the European continental shelf as determined from environmental variables: I. Tidal location method. *Mar. Biol.* 142:601–609.
- Inagake, D., Yanada, H., Segawa, K., Okazaki, M., Nitta, A. and Itoh, T. (2001) Migration of young Bluefin Tuna, *Thunnus orientalis* Temminck et Schlegel, through archival tagging experiments and its relation with oceanographic conditions in the western North Pacific. *Bull. Nat. Res. Inst. Far Seas Fish.* 38:53–81.
- Jonsen, I.D., Myers, R.A. and Flemming, J.M. (2003) Meta-analysis of animal movement using state-space models. *Ecology* 84:3055–3063.
- Kalman, R.E. (1960) A new approach to linear filtering and prediction problems. *Trans. ASME J. Basic Eng.* 82:35–45.
- Loader, C. (1999) *Local Regression and Likelihood*. New York: Springer.
- Metcalfe, J.D. (2001) Summary report of a workshop on daylight measurements for geolocation in animal telemetry. In: *Electronic Tagging and Tracking In Marine Fisheries Reviews: Methods and Technologies in Fish Biology and Fisheries*. J. Sibert & J. Nielsen (eds) Dordrecht: Kluwer Academic Press, pp. 443–456.
- Musyl, M.K., Brill, R.W., Curran, D.S. *et al.* (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*. J.R. Sibert & J.L. Nielsen (eds) Dordrecht: Kluwer Academic Press, pp. 343–368.
- 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.
- Nielsen, A. and Sibert, J. (2005) KFSST: an R-package to efficiently estimate the most probable track from light-based longitude, latitude and SST. Available at <http://www.soest.hawaii.edu/tag-data/tracking/kfst>.
- Reynolds, R.W. and Smith, T.M. (1994) Improved global sea surface temperature analyses using optimal interpolation. *J. Clim.* 7:929–948.
- 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 Technical Memorandum, NOAA-TM-NMFS-SWFC-60, 13 pp.
- Takahashi, M., Okamura, H., Yokawa, K. and Okazaki, M. (2003) Swimming behaviour and migration of a swordfish recorded by an archival tag. *Mar. Freshw. Res.* 54:527–553.
- Teo, S.L.H., Boustany, A., Blackwell, S., Walli, A., Weng, K.C. and 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.
- Vazquez, J., Perry, K. and Kilpatrick, K. (1998) *NOAA/NASA AVHRR Oceans Pathfinder Sea Surface Temperature Data Set User's Reference Manual*, Version 4.0. 10 April 1998. Pasadena, CA: Jet Propulsion Laboratory. JPL Publication D-14070.
- Wilson, R.P., Ducamp, J.-J., Rees, W.G., Culik, B.M. and Neikamp, K. (1992) Estimation of location: global coverage using light intensity. In: *Wildlife Telemetry: Remote Monitoring and Tracking of Animals*. I.G. Priede & S.M. Swift (eds). New York: Ellis Horwood, pp. 131–134.
- 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.