Horizontal movements of bigeye tuna (*Thunnus obesus*) near Hawaii determined by Kalman filter analysis of archival tagging data

JOHN R. SIBERT,[1,*] MICHAEL K. MUSYL[1] AND RICHARD W. BRILL[2]

[1]*Joint Institute of Marine and Atmospheric Research, University of Hawai'i at Manoa, Honolulu, HI 96822, USA*
[2]*National Marine Fisheries Service, Honolulu Lab/Southwest Fisheries Science Center, 2570 Dole St., Honolulu, HI 96822, USA*

ABSTRACT

Geolocation data were recovered from archival tags applied to bigeye tuna near Hawaii. A state-space Kalman filter statistical model was used to estimate geolocation errors, movement parameters, and most probable tracks from the recovered data. Standard deviation estimates ranged from 0.5° to 4.4° latitude and from 0.2° to 1.6° longitude. Bias estimates ranged from −1.9° to 4.1° latitude and from −0.5° to 3.0° longitude. Estimates of directed movement were close to zero for most fish reaching a maximum magnitude of 5.3 nm day^{-1} for the one fish that moved away from its release site. Diffusivity estimates were also low, ranging from near zero to 1000 nm^2 day^{-1}. Low values of the estimated movement parameters are consistent with the restricted scale of the observed movement and the apparent fidelity of bigeye to geographical points of attraction. Inclusion of a time-dependent model of the variance in geolocation estimates reduced the variability of latitude estimates. The state-space Kalman filter model appears to provide realistic estimates of *in situ* geolocation errors and movement parameters, provides a means to avoid indeterminate latitude estimates during equinoxes, and is a potential bridge between analyses of individual and population movements.

Key words: archival tags, geolocation errors, movement models, tuna distribution

INTRODUCTION

Archival tags are electronic devices applied to fish that record measurements of ambient light intensity and other variables at specified sampling frequencies (see Arnold and Dewar, 2001 for a review). The record of light intensity over time can be used to compute and estimate the geographical position of the tagged fish for each day the fish was at liberty. Archival tags have been successfully used in the southern Atlantic and Pacific Oceans to elucidate large-scale features of tuna movements (Anon., 1999; Lutcavage *et al.*, 1999; Block *et al.*, 2001). In geolocating archival tags, longitude is estimated from the time of local noon and latitude from the length of the day. The relationship between day-length and latitude is strongest at high latitudes and around the time of the solstices. At low latitudes and around the time of the equinoxes, this relationship weakens, and the day-length is nearly constant making the functional relationship between day-length and latitude practically indeterminate. These constraints are exacerbated by attempting to estimate the time of local noon and day-length from a light sensor attached to a fish moving freely in the water column. Thus, geolocation by light intensity has severe inherent limits to accuracy (Metcalfe, 2001). Hill and Braun (2001) suggest that under optimum conditions, the minimum error in estimating latitude is around 1° depending on the time of year and true latitude. Musyl *et al.* (2001) show that under field conditions near the tropics, errors can be much larger. Sibert and Fournier (2001) propose a method based on the Kalman filter to analyse position estimates from tracking data that estimates a 'most probable' track, geolocation errors, and parameters relevant to models of population movement. The purpose of this paper is to explore the horizontal movements of bigeye tuna (*Thunnus obesus*) near the Hawaiian Islands at 17°N latitude by applying a state-space Kalman filter statistical model to position estimates derived from archival tags.

**Correspondence*. e-mail: sibert@hawaii.edu
Received 19 September 2001
Revision version accepted 2 July 2002

METHODS

Archival tags were deployed on bigeye tuna near the Hawaiian archipelago in 1998 and 1999. Data recovered from Wildlife Computer tags were processed using the dawn and dusk symmetry method (Hill and Braun, 2001) to estimate longitude and latitude as reported by Musyl *et al.* (2001). Longitude and latitude were estimated from data recovered from the Northwest Marine Technology tag using a method that depends on the crepuscular diving behaviour of bigeye tuna. Details of tagging and data processing procedures are described in Musyl *et al.* (2003).

In addition, several tags were deployed on a mooring located at 166°42′W, 24°00′N from 26 August 1998 to 16 August 1999 (Musyl *et al.*, 2001). Data from one of these tags (number 218), a Wildlife Computer MK7 archival tag deployed at a depth of 60 m, was used to evaluate the ability of the Kalman filter to estimate geolocation errors and to produce credible position estimates.

The following presentation of a conditionally Gaussian adaptation of the extended Kalman filter statistical model for application to archival tags is based on Harvey (1990). The state-space Kalman filter model presumes a system that cannot be directly observed, but can be completely described by a vector of attributes. The Kalman filter is a statistical model consisting of an equation describing the transition of the system from one state to the next, an equation describing the errors in the process of measuring the state of the system, and a set of recursive relations that update the estimated state of the system and the components of variance at each step. The Kalman filter has found wide application in a variety of navigation-related problems (Maybeck, 1979).

For the purposes of describing the movement of a tagged fish, the state of the system at any given time is completely described by the position of the fish in two-dimensional Cartesian coordinates. The transition equation of the Kalman filter describing the movement of a fish from one time to the next is assumed to be a biased random walk on a plane:

$$\boldsymbol{\alpha}_i = \boldsymbol{\alpha}_{i-1} + \mathbf{c}_i + \boldsymbol{\eta}_i, \quad i = 1, \ldots, T \qquad (1)$$

where $\boldsymbol{\alpha}_i$ is a two-dimensional vector containing the position of the fish at time t_i, $\mathbf{c}_i$ is a two-dimensional vector representing the bias of the random walk, $\boldsymbol{\eta}_i$ is a two-dimensional vector of serially uncorrelated random variables with mean $\mathbf{0}$ and 2×2 covariance matrix, $\mathbf{Q}_i$, and T is the number of points in the time series.

Advection–diffusion models are convenient tools for describing the movements of populations of fish and in finding increasing application to the analysis of large-scale movement and distribution of tuna populations (Bertignac *et al.*, 1998; Lehodey *et al.*, 1998; Maury and Gascuel, 1999; Sibert *et al.*, 1999). Therefore, a means to estimate parameters of population movement from observations of individual fish, for example, from acoustic telemetry or from archival tags, is essential if such data are to find direct application in any spatially structured model of fisheries population dynamics. Okubo (1980) shows that the advection–diffusion equation is the continuous case of a biased random walk and that if individual movements can be described by a biased random walk, population movements can be described by the advection-diffusion equation. The advection–diffusion equation is closely related to the normal probability distribution. Feller (1966, 1968) shows that if animals are dispersing according to an advection–diffusion process, the probability of observing an animal at point x at time t is given as a normal probability density function:

$$p(t, x) = \frac{1}{\sqrt{4\pi Dt}} \exp\left(-\frac{1}{2}\frac{(x - ut)^2}{2Dt}\right). \qquad (2)$$

After a period of time t, the mean position of the animals will be given by ut and the variance will be $2Dt$. Thus, u is the mean rate of displacement, and D is the rate at which the uncertainty of the position increases over time.

These relationships provide the means to express the transition equation (1) describing individual movement by parameters applicable to describing population movement:

$$\mathbf{c}_i = \begin{pmatrix} u\Delta t \\ v\Delta t \end{pmatrix} \quad \text{and} \quad \mathbf{Q}_i = \begin{pmatrix} 2D\Delta t & 0 \\ 0 & 2D\Delta t \end{pmatrix}, \qquad (3)$$

where $\Delta t = t_i - t_{i-1}$, and u, v and D are the parameters of the advection–diffusion equation. The variables $\mathbf{c}_i$ and $\mathbf{Q}_i$ implicitly constrain the magnitude of the daily movements of the tagged fish.

The measurement equation of the Kalman filter is:

$$\mathbf{y}_i = \mathbf{Z}_i\boldsymbol{\alpha}_i + \mathbf{d}_i + \boldsymbol{\epsilon}_i, \quad i = 1, \ldots, T \qquad (4)$$

where $\mathbf{y}_i$ is a two-dimensional vector representing longitude and latitude of the fish at time t_i estimated from data recovered from an archival tag, and $\boldsymbol{\alpha}_i$ is its true (but unknown) position as described in eqn. 1. The sequence of points $\mathbf{y}_i$; $i = 1, 2, \ldots, T$ is referred below as the 'nominal' track where $\mathbf{d}_i$ is a two-dimensional vector of the bias in observing the position, and $\boldsymbol{\epsilon}_i$ a serially uncorrelated two-dimensional random vector with mean 0 and 2×2 covariance matrix, $\mathbf{H}_i$:

$$\mathbf{d}_i = \begin{pmatrix} b_x \\ b_y \end{pmatrix} \quad \text{and} \quad \mathbf{H}_i = \begin{pmatrix} \sigma_x^2 & 0 \\ 0 & \sigma_{yi}^2 \end{pmatrix}. \qquad (5)$$

The mean-squared errors in estimating longitude and latitude are σ_x^2 and σ_{yi}^2, respectively. In this coordinate system, the estimated position of the tag $\mathbf{y}_i$ is expressed in degrees of longitude and latitude, b_x and σ_x are expressed in degrees of longitude, and b_y and σ_{yi} are expressed in degrees of latitude.

b_x and b_y are equivalent to the mean 'raw' error, and σ_x and σ_{yi} are equivalent to the standard deviation of the 'raw' error reported in Musyl *et al.* (2001). Day length is a function of season and latitude. During the equinoxes, the day length is constant for all latitudes and errors in determining latitude become very large. Therefore, σ_{yi} may be allowed to vary with time (see eqn. 14 below). The (time-varying) latitude errors are symmetric. The latitude bias, b_y, is assumed to be constant over time to avoid confounding with σ_{yi}. Seasonal variability in day-length has less impact on estimation of longitude and errors in estimation of longitude are largely attributable to errors in the clock. Therefore, σ_x and b_x are assumed to be constant over time.

$\mathbf{Z}_i$ is a 2×2 matrix which changes between coordinates on the plane expressed in nautical miles (nm) and coordinates on the sphere expressed in degrees of longitude and latitude. Let $\mathbf{a}_{i|i-1}$ denote the optimal estimator of $\boldsymbol{\alpha}_i$ conditioned on all the observations up to and including $\mathbf{y}_{i-1}$ as defined below in eqn. 7:

$$\mathbf{Z}_i^{-1} = \begin{pmatrix} 60\cos(a_{i|i-1,2}/60) & 0 \\ 0 & 60 \end{pmatrix} \qquad (6)$$

where $\mathbf{a}_{i|i-1,2}$ is the estimated position of the tag north of the equator in nm at the beginning of time step *i*. The constant 60 is the number of nm per degree of latitude and per degree of longitude at the equator.

The core of the Kalman filter is a set of recursive relations that update the estimated position of the tagged fish and the components of the variance of the estimated position at each time step (Harvey, 1990). For i=1, 2,..., T,

$$\mathbf{a}_{i|i-1} = \mathbf{a}_{i-1} + \mathbf{c}_i \qquad (7)$$

$$\mathbf{P}_{i|i-1} = \mathbf{P}_{i-1} + \mathbf{Q}_i \qquad (8)$$

$$\mathbf{F}_i = \mathbf{Z}_i\mathbf{P}_{i|i-1}\mathbf{Z}_i' + \mathbf{H}_i \qquad (9)$$

$$\tilde{\mathbf{y}}_i = \mathbf{Z}_i\mathbf{a}_{i|i-1} + \mathbf{d}_i \qquad (10)$$

$$\mathbf{w}_i = \mathbf{y}_i - \tilde{\mathbf{y}}_i \qquad (11)$$

$$\mathbf{a}_i = \mathbf{a}_{i|i-1} + \mathbf{P}_{i|i-1}\mathbf{Z}_i'\mathbf{F}_i^{-1}\mathbf{w}_i \qquad (12)$$

$$\mathbf{P}_i = \mathbf{P}_{i|i-1} - \mathbf{P}_{i|i-1}\mathbf{Z}_i'\mathbf{F}_i^{-1}\mathbf{Z}_i\mathbf{P}_{i|i-1} \qquad (13)$$

Equation (7) computes the position of the tagged fish estimated from the random walk specified by the transition equation (eqn. 1). $\mathbf{a}_{i|i-1}$ can be interpreted as an estimate of the 'true' position of the tagged fish, $\boldsymbol{\alpha}_i$, at time t_i prior to making the ith observation (Maybeck, 1979). Equation (8) updates the corresponding variance of that position using the covariance matrix of the random walk, $\mathbf{Q}_i$. Equation (9) computes the total variance by combining the variance from the random walk, $\mathbf{Q}_i$, with the variance of the observation, $\mathbf{H}_i$. This variance is used in the likelihood function (15). Equation (10) computes the position estimated by the tag as specified by the measurement equation (4). Equation (11) computes the residual as the difference between the random walk position and the position estimated by the tag. This residual is used in the likelihood function (15). Equation (12) computes the most probable position as a tradeoff between the random walk position and the position estimated by the tag based on the relative variance of the two estimates. The sequence of points $\mathbf{a}_i$; $i = 1, 2, \ldots, T$ is referred below as the 'most probable' track. Equation (13) updates the variance of the most probable position using a similar tradeoff. At time t_0, it is assumed that the true position of the fish is known without error; $\mathbf{a}_0 = \mathbf{Z}_0^{-1}\mathbf{y}_0$, and $\mathbf{P}_0 = 0$, where $\mathbf{y}_0$ is the known release position. This assumption is usually well-satisfied in practice as release positions are determined using the global position system.

$\mathbf{Z}_i$ is a function of the random variable, $\mathbf{a}_{i|i-1}$, and eqn. (9) is a simple approximation of the total variance. The approximation appears to be sufficiently accurate in cases where the displacement during a single time step is relatively small, but alternative approximations for the variance of a product of random variables could be explored for cases where the displacement is large or where the positions are missing from the track.

Longitude estimation error is assumed to be constant for the duration of the time at liberty. Three different assumptions about variability in the latitude estimation error over time are considered:

$$\sigma_{yi}^2 = \begin{cases} \sigma_{y_0}^2 & 14.1 \\ \sigma_{y_0}^2(1/\cos^2\theta_i) & 14.2 \\ \sigma_{y_0}^2 e^{\xi_i} & 14.3 \end{cases} \qquad (14)$$

where $\sigma^2_{y_0}$ is the average latitude geolocation error. In case 14.1, σ^2_{yt} is assumed not to vary over time. In case 14.2, σ^2_{yt} is assumed to be functionally related to the day of the year, where $\theta_i = 2\pi(J_i+b_0)/365.25$, J_i is the number of days since the first solstice prior to the start of the track, and b_0 is a parameter to be estimated expressing the number of days prior to the equinox where the latitude error is maximal. In case 14.3, σ^2_{yt} is assumed to have a different value at each time step, and ξ_i are normally distributed random variables with mean 0 and variance σ^2_ξ representing transient deviations in the latitude estimation error.

The parameters to be estimated are u, v, D, b_x, b_y, σ^2_x, $\sigma^2_{y_0}$, and either b_0 or ξ_i depending on the latitude estimation error model. The estimates of these parameters are the values that maximize the log-likelihood function:

$$\ln L = -T \ln 2\pi - 0.5\sum_{i=1}^{T} \ln |\mathbf{F}_i| - 0.5\sum_{i=1}^{T} \mathbf{w}_i'\mathbf{F}_i^{-1}\mathbf{w}_i - z_\xi \sum_t \xi_i^2 \quad (15)$$

where z_ξ is set to reflect prior assumptions about the variance of ξ_t. The maximum of eqn. 15 is found by using a numerical procedure to find the minimum of $-\ln L$. A decrease in $-\ln L$ is equivalent to an increase in likelihood and indicates a 'more probable' track. The state-space Kalman filter model and the likelihood equation (15) were implemented in AD-Model Builder (Otter Research Ltd, 1994–1999). Upper and lower bounds were set for all parameters (Table 1) reflecting prior assumptions about the range of plausible values and previous estimates of comparable parameters (Sibert *et al.*, 1999). In practice, these bounds have no influence on the parameter estimates.

Table 1. Lower and upper bounds for parameter estimation.

Parameter	Lower	Upper	Units
u	−50.0	50.0	nm day^{-1}
v	−50.0	50.0	nm day^{-1}
D	0.0	5000.0	nm^2 day^{-1}
σ_x	0.0	15.0	degrees
σ_y	0.0	15.0	degrees
b_x	−15.0	15.0	degrees
b_y	−15.0	15.0	degrees
b_0	−50.0	50.0	days
ξ_i	−500.0	500.0	degrees

RESULTS

Tag 281 was attached to a mooring and not expected to exhibit any net displacement. Therefore, the values of u and v were held constant at zero in the analysis. The values of the estimated parameters for four different versions of the Kalman filter, the three error models given in eqn (14) and error model 14.2 with D constrained at zero, are presented in Table 2. The model with no correction for equinox error (14.1) produces estimates of latitude bias and error of −0.32 and 3.08°, respectively, nearly identical to the estimates of the standard deviation of the 'raw' error published for this tag by Musyl *et al.* (2001). The Kalman filter estimates of longitude bias and error are of −0.13 and 0.16°. The estimate of D is low, 34 nm^2 day^{-1}. Introducing the model 14.2 equinox error correction produces a more probable track, reducing the value of the negative loglikelihood function from 494 to 383, a lower estimate of D (0.19 nm^2 day^{-1}), and a correspondingly large reduction in the magnitude of the latitude errors. The maximum latitude error occurs around 10 days prior to the equinox (b_0 = 10.08) as suggested by Hill and Braun (2001). Constraining the value of D to zero has little effect on the likelihood or on the values of the parameter estimates. Application of the daily deviation equinox error model 14.3 produces a large reduction in the log-likelihood value but has little effect on the parameter estimates. Figure 1 shows the nominal and most probable 'tracks' of the mooring estimated by the Kalman filter using the model 14.2 equinox correction.

Tags were recovered from seven fish that were at liberty for brief periods of time or from which data from only a few days could be recovered. Fourteen fish were tagged with Wildlife Computer MK7 archival tags on 16 April 1999 at NOAA data buoy 51003 (Buoy 3) located at 19.17°N 160.73°W. (see Fig. 3 for positions of geographical features). Five of these fish (numbers 298, 301, 388, 390 and 392) were recaptured after short times at liberty between 19 and 35 days. A sixth fish (fish 309) was recaptured at Cross Seamount after over 1 y at liberty, but only 21 days of geolocation data were recovered from the tag. Fish 224 was tagged at Cross Seamount on 20 November 1999 and recaptured at the same location after 9 days at liberty. The effective times at liberty for these tags were very brief and most were recaptured near the release site. u and v constrained to zero for fish that were recaptured at or near the release site after short times at liberty, because there was no apparent bias in their movement. The constant latitude variance model, 14.1, was used

Table 2. Parameter estimates for all tags. Blank spaces indicates models in which the parameters were set to zero, i.e. have no influence on the model, and were not estimated. T indicates the number of points in each track. T may be less than or greater than the days at liberty if position estimates not available for all days or if the recapture position is reliable. 'Average' is the average of geolocation error estimates for the six tags released at NOAA Data Buoy 51003 and recaptured after brief times at liberty. For comparative purposes, geolocation error estimates are also shown for moored tag 218 (Musyl *et al.*, 2001), from two experiments with caged fish tunas (Gunn *et al.*, 1994), and from tests on vessels at sea (Welch and Eveson, 1999). u and v are expressed in nm day^{-1}, D in nm^2 day^{-1}, b_x, b_y, σ_x and σ_{y0} in degrees, and b_0 in days.

Tag (T)	Model	u	v	D	b_x	b_y	σ_x	σ_{y0}	b_0
Moored tag									
218 (209)	14.1			34.36	−0.13	−0.32	0.16	3.08	
218 (209)	14.2			0.19	−0.05	0.70	0.22	0.65	10.08
218 (209)	14.2				−0.01	0.76	0.23	0.65	10.08
218 (208)	14.3				−0.01	0.99	0.23	0.90	
Musyl *et al.* (2001)					−0.03	−0.40	0.24	3.08	
Tags released at Buoy 51003									
298 (24)	14.1			0.00	−0.15	−0.42	0.41	1.50	
301 (33)	14.1			0.00	−0.15	−0.15	0.66	1.81	
309 (21)	14.1			0.00	0.06	0.10	0.22	0.95	
388 (32)	14.1			0.00	−0.15	−1.92	0.46	1.03	
390 (20)	14.1			8.45	0.05	0.66	0.31	1.07	
392 (36)	14.1			23.73	−0.17	0.32	0.39	1.29	
Average	14.1				−0.08	−0.23	0.41	1.28	
Tags released at Cross Seamount									
224 (10)	14.1				−0.50	1.19	0.77	2.09	
241 (76)	14.1	5.48	−7.18	638.54	2.89	4.06	0.34	7.56	
241 (76)	14.2	5.31	−4.40	333.74	2.98	2.59	0.43	0.49	9.90
241 (76)	14.2	7.71	−2.48	1049.73			0.19	0.45	9.91
241 (76)	14.3	5.29	−4.24	323.91	2.99	2.48	0.43	3.49	
625 (52)	14.1			0.00	−0.31	2.06	0.57	2.49	
Tag released at Kona Coast of Island of Hawaii									
509 (80)	14.2			25.60	−1.24	−1.49	1.56	2.52	2.96
Comparable studies									
Gunn *et al.* (1994)					0.54	1.92	0.09	0.22	
Welch and Eveson (1999)					2.0	1.3	0.9	1.2	

for all tags with short times at liberty. The parameter estimates from the Kalman filter are tabulated in Table 2. Diffusion estimates are very low, ranging from near zero to a maximum of 24 nm^2 day^{-1}. The average longitude geolocation error (σ_x) for the Buoy 3 releases is 0.41°; the average latitude geolocation error (σ_{y_0}) is larger, approximately 1.28°. These parameter estimates are consistent with estimates from the moored tag and can be considered as benchmarks for fish that did not stray far from their release site. Attempts were made to estimate u and v from the mooring data, but the numerical minimization of the likelihood function (15) did not converge to a solution.

Fish 241 was tagged with a Wildlife Computer MK7 archival tag on 21 January 1999 at Cross Seamount (18.48°N, 158.25°W) and recaptured 20 October 1999 in the same vicinity. Only the first 77 days of data were recovered from the total of 272 days the fish was at liberty, a period spanning the 1999 vernal equinox, and the recapture day was not included in the recovered data. Table 2 shows the results of the Kalman filter applied to the nominal geolocation estimates computed from the data recovered from the tag using three different models of latitude estimation error. The uniform error model, 14.1, produces a very large estimate of the latitude error, $\sigma_{y_0} = 7.56°$. Introduction of the cosine model, 14.2, produces a much smaller estimate of $\sigma_{y_0} = 0.49°$. The daily deviation model, 14.3, with $z_\xi = 1.0$ produces an intermediate value of $\sigma_{y_0} = 3.49°$. The latitude geoposition error for archival tags is known to be very large; setting $z_\xi = 1.0$ implies that coefficient of variation of the latitude

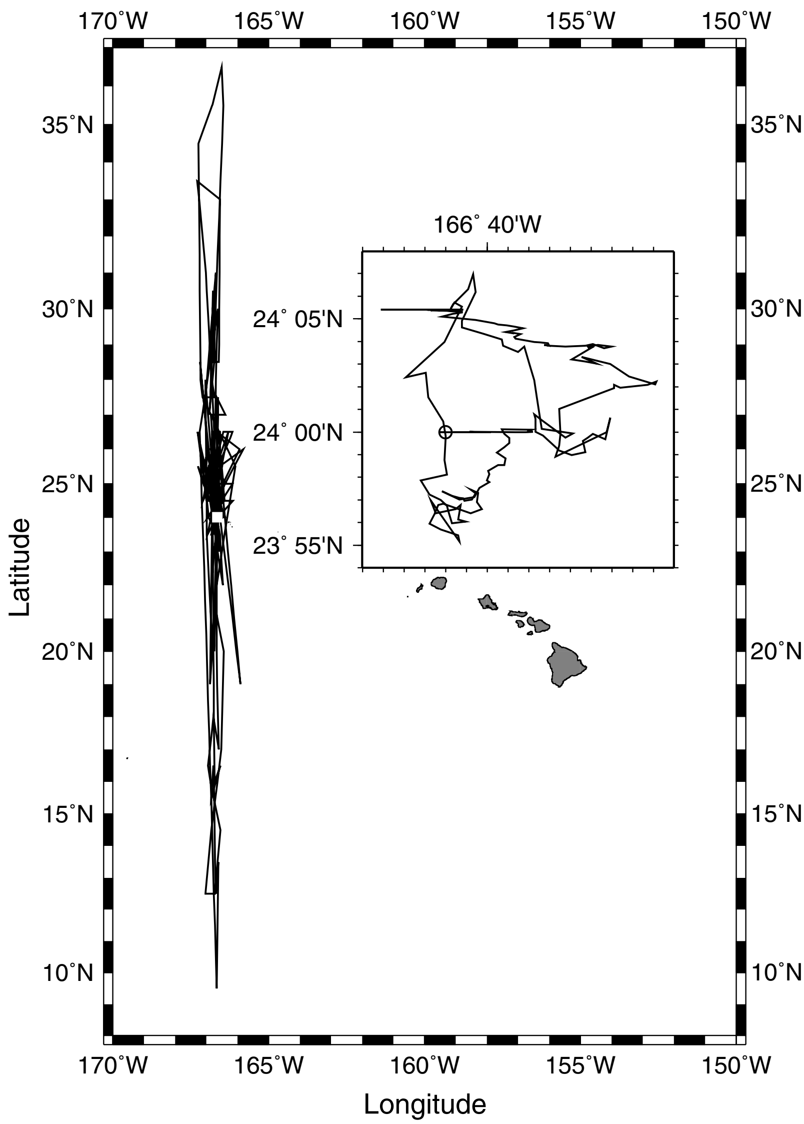

Figure 1. Nominal 'track' of tag 218 deployed on a mooring at 166°42′W, 24°00′N estimated by the geolocation algorithm. Open square in track indicates the approximate location of the mooring and area shown in the inset. Inset map is an enlargement of the open square and shows the most probable 'track' from the six-parameter model with $D = 0.19$ nm^2 day^{-1} and latitude error model 14.2. The open circle in the inset indicates the position of the mooring.

estimates is approximately 0.7. Figure 2 illustrates the differences in the three error models. The cosine equinox error model becomes very large 10 days prior to the equinox. The daily deviation model generally tracks the cosine error model but is more irregular, as might be expected. Similar trends were observed for moored tag 218.

Figure 3 shows the most probable track for fish 241 using the cosine equinox error model. This fish appeared to pass the first part of its time at liberty near Cross Seamount and Buoy 2 before making a move towards the east.

Fish 509 was tagged with Northwest Marine Technology archival tag on 9 April 1998 near the Kona coast of the Island of Hawaii. It was recovered on 2 July 1998, 84 days later, in the same general vicinity. This fish measured 131 cm in fork length and was the largest fish recaptured. The Kalman filter was applied to the geolocation estimates from the tag data with u and v constrained to zero (Table 2). The estimates of D are fairly low. Addition of the seasonal latitude error model results in a decrease in the log-likelihood value and a corresponding decrease in the estimate of σ_{y_0}. The estimated most probable track is shown in Fig. 4. It appears to have remained near the south end of the Island of Hawaii.

Fish 625 tagged with a Wildlife Computer MK7 archival tag was placed on a juvenile (65 cm) bigeye tuna at Cross Seamount on 19 November 1999 and recaptured near the same area on 9 January 2000 after 51 days at liberty. The Kalman filter parameter estimates are presented in Table 2. The estimate of D is

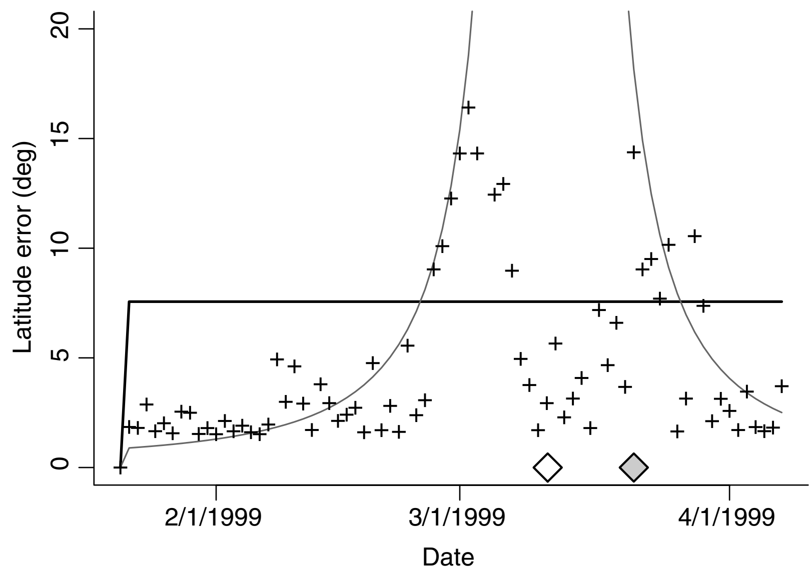

Figure 2. Estimates of latitude geolocation error using three variance models for fish 241. The heavy line is the uniform error model (14.1), the light line is the seasonal error model (14.2), and the crosses are daily deviation model (14.3). The open diamond on the abscissa indicates the position of the maximum of the seasonal error model at $b_0 = 9.90$ days prior to the equinox. The shaded diamond on the abscissa indicates the date of the equinox.

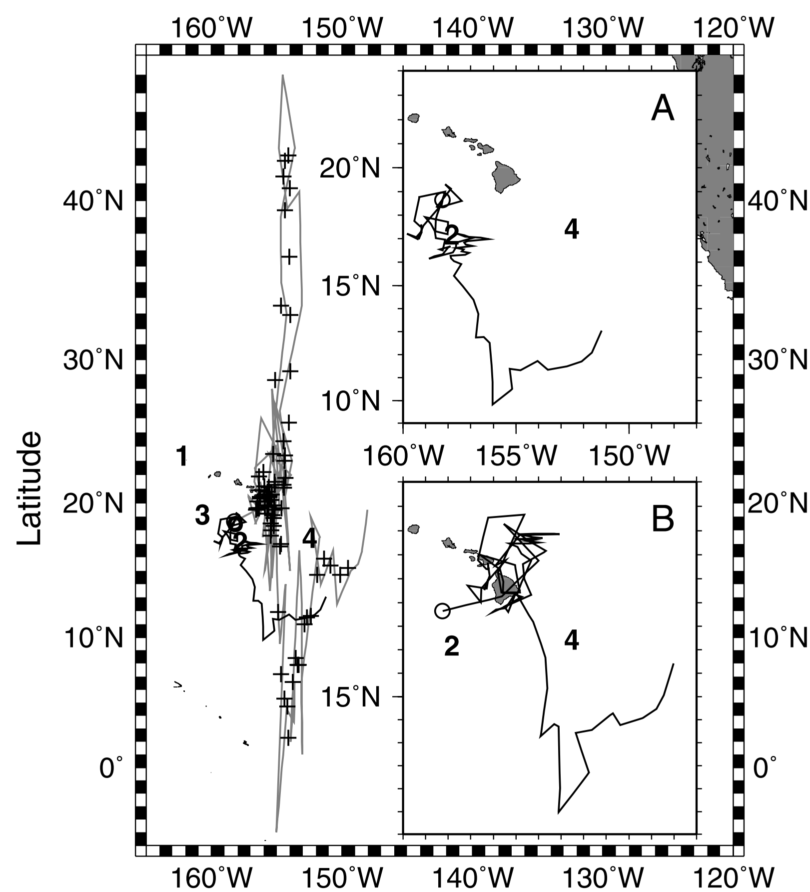

Figure 3. Nominal track of fish 241 estimated by the geolocation algorithm (light line) and most probable track estimated by the Kalman filter using equinox error model 14.2 (dark line). The symbols '1', '2', '3' and '4' indicate positions of NOAA Data Buoys 51001, 51002, 51003 and 51004, respectively. The open circle indicates the release position at Cross Seamount. Crosses are positions computed by a five-point moving average of the nominal positions. Inset map A is an enlargement of the most probable track with b_x and b_y estimated as shown in Table 2. Inset map B is an enlargement of the most probable track with b_x and b_y constrained at zero.

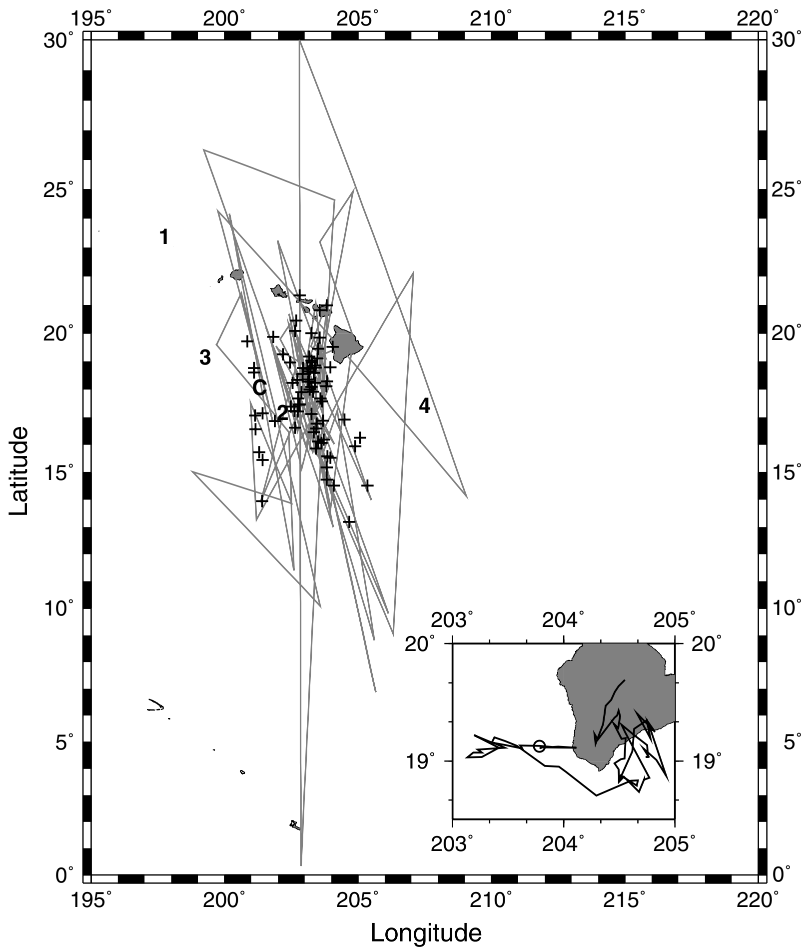

Figure 4. Nominal track of fish 509 estimated by the geolocation algorithm (light line). Crosses are positions computed by a five-point moving average of the nominal positions. The symbols '1', '2', '3' ,'4' and 'C' indicate positions of NOAA Data Buoys 51001, 51002, 51003, 51004, and Cross Seamount, respectively. The inset map shows the region near the Kona Coast of the Island of Hawaii and the most probable track estimated by the Kalman filter using the six-parameter model with equinox error 14.2; the open circle indicates the release position.

close to zero and constraining u and v has little impact on either the value of the log likelihood function or on the estimate of D.

DISCUSSION

The inherent limits to accuracy of geolocation by light intensity are exacerbated around the time of the equinoxes. Examination of the tracks in Figs 3 and 4 indicates that simply smoothing the tracks with a moving average does not produce a less variable or more credible series of positions. The Kalman filter is not a smoothing procedure. Rather, the Kalman filter computes the position of the fish as a weighted average of two position estimates, the 'nominal' geolocation position and the position computed by the transition equation, with weights inversely proportional to the current estimate of the variance ($\mathbf{P}_{i|i-1}$) and the measurement error ($\mathbf{H}_i$). Time-dependent models of the latitude geolocation error, 14.2 and 14.3, amplify latitude estimation error in the measurement equation (4). The Kalman filter, therefore, attributes more credibility to the transition equation estimate and less credibility to the geolocation estimate. The variance weighting and the displacement constraints implicit in the transition equation enable the estimation of plausible daily positions during periods of time when geolocation algorithms are least reliable, alleviating one of the most vexing shortcomings of archival tags.

Both the cosine and the daily deviation models of latitude error appear to reduce the average latitude geolocation error. In the cases presented here, the cosine-error model produces slightly lower geolocation error estimates, but in the absence of more examples, it is difficult to find an empirically based preference for one over the other. However, the cosine-error model mimics more closely the day-length around the equinox by introducing a singularity into the model.

Smith and Goodman (1986) show that ocean temperature information can be used to estimate latitude, assuming that longitude is known without error, in areas where there is a sufficient temperature gradient. They calculate that a resolution as low as 2° of latitude could be achieved using sea surface temperature. Several authors (e.g. Block *et al.*, 2001) have successfully used ocean temperature fields in conjunction with temperature measurements from archival tags to estimate or confirm latitude for long-distance movements. There was insufficient contrast in the available temperature fields around the Island of Hawaii to apply this method to the data from the bigeye tags. The Kalman filter produces much smaller latitude errors than those anticipated by Smith and Goodman, even during periods of the equinox. The examples reported here indicate that the method works well at low latitudes, and one would expect it to be equally effective at higher latitudes where day-length differences are more pronounced.

Geolocation error estimates produced by the Kalman filter applied to moored tag 218 are nearly identical to estimates of bias (average 'raw' error) and standard deviation (standard deviation of the 'raw' error) reported by Musyl *et al.* (2001) for this tag. Geolocation errors for archival tags have been estimated by deployment on fish confined to cages (Gunn *et al.* 1994), on moorings, and on moving vessels (Welch and Eveson, 1999). The error estimates produced by the Kalman filter applied to data from archival tags deployed on free roaming fish are comparable with estimates obtained under controlled conditions (Table 2).

The longitude bias (b_x) estimates for two of the tags (fish 241 and 509) are larger than expected, 2.98° and −1.24°, respectively, in comparison with estimates of b_x close to zero from the mooring experiment and the buoy 3 releases. Inspection of the position estimates from the geolocation algorithm reveals that the first position estimates are about 2.8° east and 0.9° west of their respective release positions. Either these fish made substantial movements in their first 24 h of liberty, release positions were not recorded accurately, or the clocks in the tags were in error. Clock errors in archival tags are computed by comparing the time of the tag with a reference time standard, such as United States Naval Observatory. Clock errors are generally very low, of the order of a few seconds, but in some instances, errors exceeding 2 h were recorded. The clock in tag 509 was in error by −4 min and 2 s, an error sufficient to explain about 1° of the estimated −1.24° of bias. To produce the 2.98° of bias estimated for tag 241, the clock should have an error by approximately 12 min. Unfortunately, a malfunction in tag 241 prevented determination of clock error so that a 12-min error cannot be excluded. Figure 3B shows the most probable track estimated with b_x and b_y constrained at zero. The track is shifted to the north-east so that the early part of the track is near the Island of Hawaii. The differences in the estimated parameters (Table 2) indicate greater mobility (higher estimates of u and D), lower longitude error and a generally more variable track. Comparison of nominal and predicted recapture positions would help resolve this discrepancy. Unfortunately, either recapture positions were not recorded accurately or the data recovered from these three tags did not include the recapture date. Ambiguities in interpretation of bias estimates could be resolved using 'pop-up' positions reported by archival tags that report stored data via satellite.

It is tempting to use the values of the likelihood function for model selection. However, eqn. 15 does not compare observed and predicted tracks as there are no observed tracks. Rather, there are two different tracks, one estimated by a geolocation algorithm and the other by the Kalman filter applied to the geolocation estimates. The track estimated by the Kalman filter is the track for which eqn. 15 is maximized and can be interpreted as the track with the least variance consistent with the presumed movement of the fish and the geolocation errors. Testing the significance of changes in $-\log L$ to discriminate between models makes the implicit assumption that the fish are moving with minimum variance, an assumption for which there is no *a priori* justification. It is always possible to achieve a reduction in $-\log L$ by constraining diffusion (D) to zero.

None of the tagged fish appears to have strayed far from the site at which it was tagged. This result supports the previous conclusion that bigeye tunas have a strong tendency to remain associated with offshore features such as weather buoys and sea mounts (Holland *et al.*, 1999; Itano and Holland, 2000). The apparent association of a large fish (number 509) with the Island of Hawaii suggests that this conclusion may also be true of adult fish and implies that points of attraction in a mid-ocean habitat are not restricted to floating objects and seamounts – larger features, such as islands, also act as points of attraction for bigeye tuna.

With the exception of fish 241, the estimated movement parameters presented in Table 2 are lower than those estimated by a large-scale, advection–diffusion model applied to conventional tagging data on tropical skipjack (*Katsuwonus pelamis*) in the western Pacific (Sibert *et al.*, 1999). Low numerical values of the movement parameters are consistent with the restricted scale of the observed movement and the apparent fidelity of bigeye to geographical attractors. Fish 241 is the only fish that recorded a net displacement, and the estimates of the movement parameters for this fish are comparable with those for skipjack in the tropics.

These results of initial attempts to apply the Kalman filter to archival tagging data of short duration are very encouraging. The Kalman filter produces estimates of geolocation errors for tags applied to fish in actual field conditions and movement parameters applicable to population-scale models. The method, in principle, can be applied to any tracking data and is not restricted to archival tags. The Kalman filter also provides plausible position estimates during the equinoxes, periods which the geolocation by light is least reliable.

The state-space Kalman filter model is easily extended. We are currently applying the model to longer tracks and developing models which share parameters among several tracks. The order of both $\boldsymbol{\alpha}$ and $\mathbf{y}$ can be increased to include ambient temperature recorded by most archival tags. Such a model would require revision of the transition equation to update the temperature from observed temperature fields. The order of $\mathbf{y}$ could also be increased to accommodate additional position estimates, such as might be produced by ARGOS tracking in double tagging experiments. A means to allow both movement and error parameters to vary over time would accommodate long tracks in which the behaviour of the animal varies, the performance of the archival tag changes with time (as might be expected if biofouling were interfering with light measurements), and for cases where the tag is at liberty through more than one equinox. Finally, further work is required on the issue of statistical model selection in order to evaluate the significance of including or excluding model parameters.

ACKNOWLEDGEMENTS

We would like to thank Roger Hill of Wildlife Computers for clarifying the fine points of geolocation by light level, Dave Fournier from Otter Research for assistance in implementing the Kalman filter in ADModel Builder, Pierre Kleiber from NMFS for suggestions regarding the Kalman filter, and Mark Maunder of the Inter-American Tropical Tuna Commission for constructive comments on the manuscript. The work described in this paper was sponsored by the University of Hawaii Pelagic Fisheries Research Program under Cooperative Agreement number NA67RJ0154 from the National Oceanic and Atmospheric Administration.

REFERENCES

Anon. (1999) Final Report. In: *Southern Bluefin Tuna Recruitment Monitoring Program (1993–1998)*. T.L.O. Davis (ed.) Hobart: CSIRO Marine Research, pp. 7–27.

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. Sibert and J. Nielsen (eds) Dordrecht: Kluwer Academic Press, pp. 7–74.

Bertignac, M., Lehodey, P. and Hampton, J. (1998) A spatial population dynamics simulation model of tropical tunas using a habitat index based on environmental parameters. *Fish. Oceanogr.* **7**:326–334.

Block, B.A., Dewar, H., Blackwell, S.B. *et al.* (2001) Archival and Pop-up satellite tagging of Atlantic Bluefin Tuna. In: *Electronic Tagging and Tracking in Marine Fisheries Reviews: Methods and Technologies in Fish Biology and Fisheries*. J. Sibert and J. Nielsen (eds) Dordrecht: Kluwer Academic Press, pp. 65–88.

Feller, W. (1966) *An Introduction to Probability Theory and its Applications*, Vol. II. John Wiley & Sons, New York.

Feller, W. (1968) *An Introduction to Probability Theory and its Applications*, Vol. I. (3rd edn). John Wiley & Sons, New York, 509 pp.

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*.

Harvey, A.C. (1990) *Forecasting, Structural Time Series Models and the Kalman Filter*. Cambridge University Press, Cambridge.

Hill, R.D. and Braun, M.J. (2001) Geolocation by light-level. The next step: latitude. In: *Electronic Tagging and Tracking in Marine Fisheries Reviews: Methods and Technologies in Fish Biology and Fisheries*. J.Sibert and J. Nielsen (eds) Dordrecht: Kluwer Academic Press, pp. 443–456.

Holland, K.N., Kleiber, P. and Kajiura, S.M. (1999) Different residence times of yellowfin tuna, *Thunnus albacares*, and bigeye tuna, *T. obesus*, found in mixed aggregations over a seamount. *Fish. Bull.* **97**:392–395.

Itano, D.G. and Holland, K.N. (2000) Movement and vulnerability of bigeye (*Thunnus obesus*) and yellowfin tuna (*Thunnus albacares*) in relation to FADs and natural aggregation points. *Aquat. Living Res.* **13**:213–223.

Lehodey, P., Andre, J., Bertignac, M. *et al.* (1998) Predicting skipjack tuna forage distributions using a coupled dynamical bio-geochemical model. *Fish. Oceanogr.* **7**:317–325.

Lutcavage, M.E., Brill, R.W., Skomal, G.B., Chase, B.C. and Howey, P.W. (1999) Results of pop-up satellite tagging of spawning size class fish in the Gulf of Maine: do North Atlantic bluefin tuna spawn in the mid-Atlantic? *Can. J. Fish. Aquat. Sci.* **56**:173–177.

Maury, O. and Gascuel, D. (1999) SHADYS, a GIS based numerical model of fisheries. Example application: the study of a marine protected area. *Aquat. Living Resour.* **12**:77–88.

Maybeck, P.S. (1979) *Stochastic Models, Estimation, and Control*, Vol. 1. Academic Press, Republished 1994 by Navtech Book and Software Store, Arlington, VA.

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 and 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. Sibert and J. 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.

Okubo, A. (1980) *Diffusion and Ecological Problems: Mathematical Models*. Springer, New York, 254 pp.

Otter Research Ltd(1994–1999). *AD Model Builder Documentation*. Otter Research Ltd, Sidney, British Columbia. http://otterrsch.com.

Sibert, J.R., Hampton, J., Fournier, D.A. and Bills, P.J. (1999) An advection–diffusion-reaction model for the estimation of fish movement parameters from tagging data, with application to skipjack tuna (*Katsuwonus pelamis*). *Can. J. Fish. Aquat. Sci.* **56:**925–938.

Sibert, J. and Fournier, D.A. (2001) Possible models for combining tracking data with conventional tagging data. In: *Electronic Tagging and Tracking in Marine Fisheries Reviews: Methods and Technologies in Fish Biology and Fisheries*, J. Sibert and J. Nielsen (eds) Dordrecht: Kluwer Academic Press, pp. 443–456.

Smith, P. and Goodman, D. (1986) *Determining fish movement from and '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.

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.