Possible Models for Combining Tracking Data with Conventional Tagging Data

John Sibert[1] and David Fournier[2]
[1] *University of Hawaii, Honolulu, e-mail jsibert@soest.hawaii.edu*
[2] *Otter Research, Sydney, British Columbia, e-mail otter@otter-rsch.com*

Abstract   Advection-diffusion models have been used successfully to describe the time and place of recapture of tuna tagged with conventional dart tags. Such models are the continuous analogs of a biased random walk. This paper demonstrates how biased random walks can be used to simulate large scale movements of tunas as recorded by archival tags in a way that captures all of the major characteristics of the tracks. The parameters of the biased random walk model are identical to the parameters of the advection diffusion model, suggesting that a joint parameter estimation procedure might be feasible. Finally, the potential application of the Kalman filter to the analysis of tracking data is discussed. This statistical model has the potential to increase the accuracy of geoposition estimates from tracking devices as well as to estimate biased random walk parameters from tracking data.

1.   INTRODUCTION

Acoustic transponders and archival tags provide large amounts of data pertaining to the vertical and horizontal position of individual fish. Ancillary sensors provide additional information such as internal body temperature and ambient water temperature. In contrast, conventional "dart" or "spaghetti" tags provide a small amounts of data pertaining to the horizontal position of a large number of fish and no data at all pertaining to their vertical position. Whether the data derived from the use of these devices are informative depends on the scientific questions being addressed and the means used to analyze the data. Tracking data on vertical distribution and apparent temperature preferenda have been particularly informative when analyzed in the context of attempting to quantitatively define the habitat of fish species in terms of depth, temperature and oxygen preferences. The results have found ready ap-

plication in the analysis of catch and effort data from longline fisheries. Indices of abundance have been refined by estimating the number of longline hooks deployed in the depth strata most likely to be frequented by the species of interest (Hinton and Nakano, 1996; Hampton et al., 1998). It remains to be seen whether the accuracy of these indices of abundance is improved.

Data pertaining to horizontal movements have proven similarly useful when analyzed in the context of determining the prefered habitat of tunas, but have been slower to penetrate the quantitative fisheries stock assessment and management mainstream. While there are undoubtedly many reasons for this lack of application, one major barrier is the lack of a generally accepted context in which to apply data on the movement of individual fish. Stock assessment is concerned with estimation of population-level processes, such as mortality, that cannot easily be estimated from a track of an individual fish. Furthermore, unknown position errors in data from archival tags makes the application of these data problematical.

In this paper, we introduce the related concepts of a biased random walk and the diffusion-advection equations as a consistent theoretical framework for combining tracking data with conventional with tag release and recapture data. A simple biased random walk simulation of a tuna track is presented to demonstrate that this framework may be suitable for analysis of tracking data. The results of a preliminary statistical model for estimating diffusion-advection parameters from long-distance tracks is then presented.

2.   BIASED RANDOM WALKS

For simplicity, the following discussion is restricted to one dimensions; the ideas are easily extended to two dimension. In a simple random walk, an animal is assumed to move a constant small distance, λ, in a constant small interval of time, τ, as shown schematically below.

$$\longleftarrow\!\ \lambda \ \longrightarrow\!\!\!\longleftarrow\!\ \lambda \ \longrightarrow$$

$x_0 - \lambda$	x_0	$x_0 + \lambda$
$t + \tau$	t	$t + \tau$
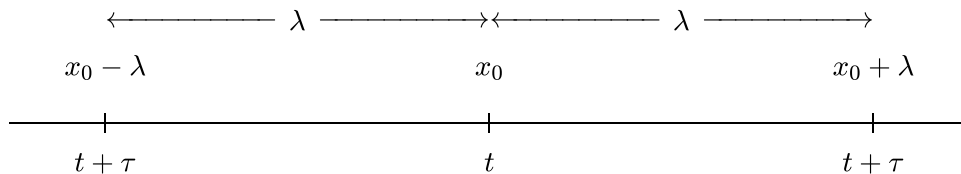
At time 0, the animal is at position x_0 and can move either to the right with probability P_1 or to the left with probability P_2. At time $t + \tau$, the animal is at position $x_0 + \lambda$ with probability P_1 or at position $x_0 - \lambda$ with probability P_2; $P_1 + P_2 = 1$. The difference in probability of moving to the left or to the right, the "bias", is $\epsilon = P_1 - P_2$. Although

a random walk is a stochastic process and the position of an individual fish cannot be predicted exactly, the fish's movement is not haphazard. The fish moves at constant speed λ/τ, but the direction of movement is not constant. Therefore it can never move farther than λ in any time step. Furthermore, it is possible to imagine that the bias or the distance moved in a time step might be variable in time and space and may be dependent on characteristics of the environment or may reflect some inherent tendency to prefer one direction over another.

The advection-diffusion equation is closely related to a biased random walk. Okubo (1980) derives the advection-diffusion equation from a random walk by showing that

$$\lim_{\lambda,\tau,\epsilon\to 0} \epsilon \cdot \frac{\lambda}{\tau} = u \qquad \lim_{\lambda,\tau\to 0} \frac{\lambda}{2} \cdot \frac{\lambda}{\tau} = D \qquad (1.1)$$

and that the rate of change in density of animals at a point can be expressed as

$$\frac{\partial N}{\partial t} = -u\frac{\partial N}{\partial x} + D\frac{\partial^2 N}{\partial x^2} \qquad (1.2)$$

In other words, the advection-diffusion equation is the limiting case of a biased random walk and there is a straightforward interpretation of the parameters of one model in the context of the other.

Advection-diffusion models are finding increasing application to the analysis of large-scale movement and distribution of tunas (Lehodey et al., 1998; Bertignac et al., 1998; Maury and Gascuel, 1999). Advection-diffusion models, in which the parameters vary regionally and seasonally, can predict the time and place of recapture of tagged skipjack tuna (Katsuwonus pelamis) with reasonable accuracy (Sibert et al., 1999). Therefore it should be possible to model the track of a singe tuna using a biased random walk parameterized to be consistent with an advection-diffusion model of a population of conventionally tagged tunas.

3.   RANDOMLY WALKING ACROSS THE PACIFIC

Tsuji et al. (1999) presented preliminary data from an archival tag recovered from a Pacific northern bluefin tuna, Thunnus thynnus. The animal was tagged in November, 1996 near Tsushima Island and recaptured in August 1998 off California. The track presented by Tsuji et al. (1999) is reproduced in Figure 1. As the authors note in their paper, the animal appeared to change its behavior several times during the period it was at liberty. Whether these changes are due to season, geography or local environmental conditions must await a more complete analysis.
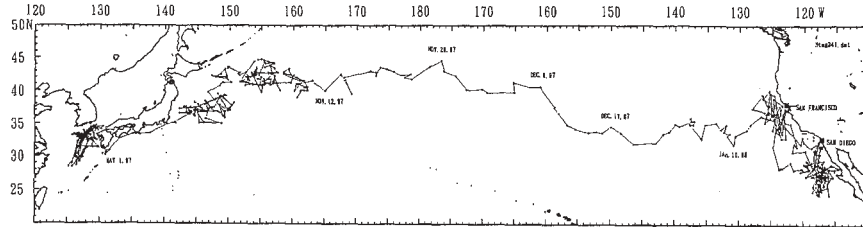
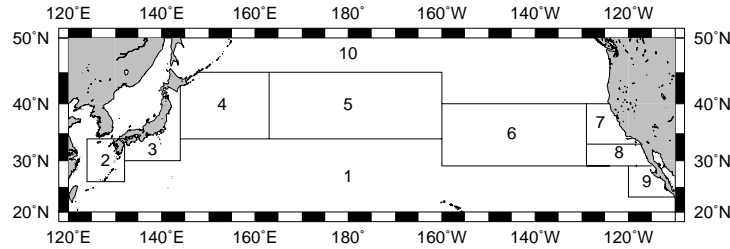*Figure 1*   Track of juvenile bluefin tuna from Tsuji et al. 1999.

Figure 2   Regions used for bluefin tuna random walk simulations.

Sibert et al. (1999) superimposed an array of regions over the model domain, in which the parameters of the advection-diffusion model were assumed constant. The same principle can be applied to the tuna track in Figure 1. Eight model regions were designated (Figure 2), corresponding to regions where the tagged bluefin appeared to have uniform behavior. Two additional model regions were added to the north of the track and the south of the track. Values of the advection-diffusion parameters, u, v, and D, were assigned to each region. A characteristic speed, $\mu = \lambda/\tau$, was assigned to be similar to daily movement speed estimated from the bigeye track, approximately 100 Nmi da^{-1}. The time step was fixed at 1 hour. Biases in the north-south (ϵ_y) and east-west (ϵ_x) dimensions were calculated from the parameters u and v using the relationships in equation (1.1). The distance moved at each time step was calculated from the parameter D and the characteristic speed, $\lambda = \frac{D}{2\mu}$. Two uniform random numbers between 0 and 1 were generated at each time step and compared with ϵ_y and ϵ_x to determine the direction of the jump in the north-south and east-west dimensions. The simulation was run for twenty months and the positions of simulated fish were recorded once per day. The numerical values of the parameters in each region are shown in Table 1.

Table 1 Advection-diffusion parameters used for biased random walk simulation of trans-Pacific bluefin tuna track. u and v in $\mathrm{Nmi\,da^{-1}}$; D in $\mathrm{Nmi^2da^{-1}}$.

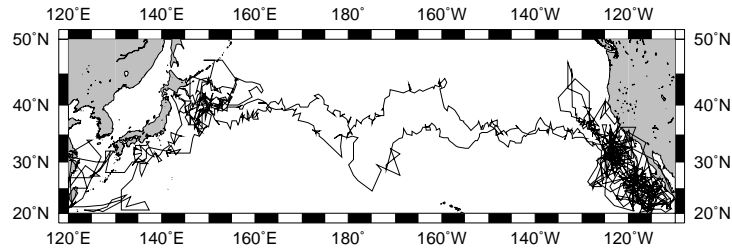| Region	u	v	D
1	2.0	10.0	100
2	4.0	3.0	100
3	10.0	6.0	33
4	2.0	2.0	33
5	15.0	0.0	23
6	15.0	0.0	23
7	0.0	0.0	100
8	0.0	0.0	167
9	0.0	0.0	100
10	0.0	0.0	100

Figure 3 Two simulated bluefin tuna tracks.

Figure 3 shows two representative tracks generated by two different series of random numbers. These two simulated tracks display all of the important features of the observed bluefin track. There are places where the simulated fish appeared to have no prefered direction of movement and places where it moved consistently towards the East. The two fish did not arrive at the West coast of North America simultaneously, and both show a "bimodal" distribution after arriving off the West Coast. These results suggest that a statistical model of a tuna track based on a biased random walk is feasible.

4. PRELIMINARY STATISTICAL MODEL

A useful statistical model for application to tracking data would estimate the parameters of a biased random walk parameterized consistently with statistical models of the advection-diffusion equation applied

to conventional tags. In addition, the model would estimate the errors in the position estimates from the tracking device so that a "corrected" track could be estimated.

The advection-diffusion equation is closely related to the normal probability distribution. Feller (1968) shows that if the probability of observing an animal at point x at time t is given as a normal probability density function,

$$p(t, x) = \frac{1}{\sqrt{4\pi Dt}} e^{-\frac{1}{2}\frac{(x-ut)^2}{2Dt}} \tag{1.3}$$

the function p uniquely satisfies equation (1.2). In other words, if animals are dispersing according to equation (1.2), after a period of time t their mean position will be given by ut and the variance of their position will be $2Dt$. Thus, u can be considered to be the mean rate of displacement (which may be zero) and D a measure of the rate at which the uncertainty of the position increases with time (Feller, 1966).

The notions of a state space model and the Kalman filter (Harvey, 1990) form the basis for a statistical model of a track. Let

$$y_t = \alpha_t + d_t + \epsilon_t, \quad t = 1, \dots, T \tag{1.4}$$

where y_t is a two dimensional vector representing the observed position of the tagged fish at time t, α_t its true position, d_t a two dimensional vector the bias in observing the position, and ϵ_t a serially uncorrelated 2 dimensional random vector with mean 0 and 2×2 covariance matrix, H_t,

$$d_t = \begin{pmatrix} 0 \\ 0 \end{pmatrix} \quad \text{and} \quad H_t = \begin{pmatrix} \sigma_x^2 & 0 \\ 0 & \sigma_y^s \end{pmatrix}. \tag{1.5}$$

In other words, we assume that there is no systematic bias in the tag's position estimate and that there are independent random errors in the estimate of longitude and latitude, where σ_x^2 and σ_y^2 are the mean square errors of the longitude and latitude estimates.

The true position of the tagged fish is not observed, but is assumed to be generated by a biased random walk

$$\alpha_t = \alpha_{t-1} + c_t + \eta_t, \quad t = 1, \dots, T \tag{1.6}$$

where c_t is a 2 dimensional vector representing the bias of the random walk, and η_t is a 2 dimensional vector of serially uncorrelated random variables with mean 0 and covariance matrix, Q_t. For the biased random walk in equation (1.6)

$$c_t = \begin{pmatrix} u \\ v \end{pmatrix} \quad \text{and} \quad Q_t = \begin{pmatrix} 2D & 0 \\ 0 & 2D \end{pmatrix}. \tag{1.7}$$

The Kalman filter consists of a set of recursive relations that update the estimated position of the tagged fish and the components of the variance of the estimated position.

$$
\begin{aligned}
a_{t|t-1} &= a_{t-1} + c_t \\
P_{t|t-1} &= P_{t-1} + Q_t \\
w_t &= y_t - a_{t|t-1} - d_t \\
F_t &= P_{t|t-1} + H_t \\
a_t &= a_{t|t-1} + P_{t|t-1} F_t^{-1} w_t \\
P_t &= P_{t|t-1} - P_{t|t-1} F_t^{-1} P_{t|t-1}
\end{aligned}
\tag{1.8}
$$

where P_t is a 2×2 covariance matrix of the estimation error and $a_{t|t-1}$ can be interpreted as an estimate of the "true" position of the tagged fish, α_t.

The parameters to be estimated are u, v, D, σ_x^2, and σ_y^2. The estimates of these parameters are the values that maximize the log likelihood function

$$
\ln L = -T \ln 2\pi - 0.5 \sum_{t=1}^{T} \ln |F_t| - 0.5 \sum_{t=1}^{T} w_t' F_t^{-1} w_t
\tag{1.9}
$$

The performance of this model was examined using a series of simulations. For each series, one thousand replicate simulated archival tag tracks were generated for assumed values of the model parameters u, v, D, σ_x^2, and σ_y^2. The RMS error of the geolocation estimates produced by currently available archival tags is approximately $\frac{1}{2}$ degree of longitude and $1\frac{1}{2}$ degrees of latitude (R. Hill, personal communication). These errors were fixed for all simulations at 30 and 90 nautical miles (Nmi) respectively assuming the tracks to be near the equator. Three behavior scenarios were simulated: (1) net eastward movement with movement variability about the same order of magnitude as geolocation error; (2) net eastward movement with movement variability about half of the geolocation error; and (3) no net movement and movement variability about half of the geolocation error. The numerical values of the input parameters are given in Table 2.

Simulation results were evaluating by computing the bias as the difference between the input value of a parameter and the value estimated by the Kalman filter. The bias was "standardized " by dividing by the standard deviation of the estimates computed over all 1000 simulations so that parameters with different ranges could be easily compared. This

Table 2 Input parameter values for simulations of Kalman filter estimation procedure. u and v in Nmi da^{-1}; D in $\text{Nmi}^2\text{da}^{-1}$; σ_x and σ_y in Nmi. The column headed 'sd' is the standard deviation of the movement uncertainty in Nmi.

Simulation	σ_x	σ_y	u	v	D	sd
1	30	90	50	0	1800.0	60
2	30	90	50	0	112.5	15
3	30	90	0	0	112.5	15

comparison evaluates the ability of the Kalman filter estimator to correctly recover the input parameters given a large number of replicate tracks.

The worst (lowest value of the likelihood function, equation (1.9)) and best (highest value of the likelihood function) fitting simulations were examined by Markov-chain Monte Carlo exploration of the likelihood surface. This analysis treats the parameter estimates as random variables and computes the posterior probability distribution of the estimates given the observations (i.e., the simulated track) and the constraints on the parameters. It evaluates the ability of the Kalman filter estimator to recover the parameters from a single track.

5.   RESULTS

Figure 4 shows "box and whisker" diagrams of the standardized bias for all three simulations. The mean bias, given by the horizontal line within the boxes, is close to zero for all variables and the estimates appear to be approximately normally distributed. The model appears to slightly overestimate diffusivity, D.

Figure 5 compares the likelihood profiles of all the parameter estimates for a "good" fit, L1754 (ln $L = -1753.94$) and a "bad" fit, L2053 (ln $L = -2053.14$). For both cases, the estimates of the random walk bias parameters u and v are well behaved, their estimates are very close to the true values, and the posterior probability of the estimates given the data appears normal. The posterior distribution of the estimates of diffusivity, D, are asymmetric with the point estimates higher than the "true" value. Nevertheless, the true value of D lies within two standard deviations of the estimated value. In the case of the "good" fit, which happens to be a model with diffusivity less than geolocation error, the true value of the longitude error is more than two standard deviations higher than the estimate.
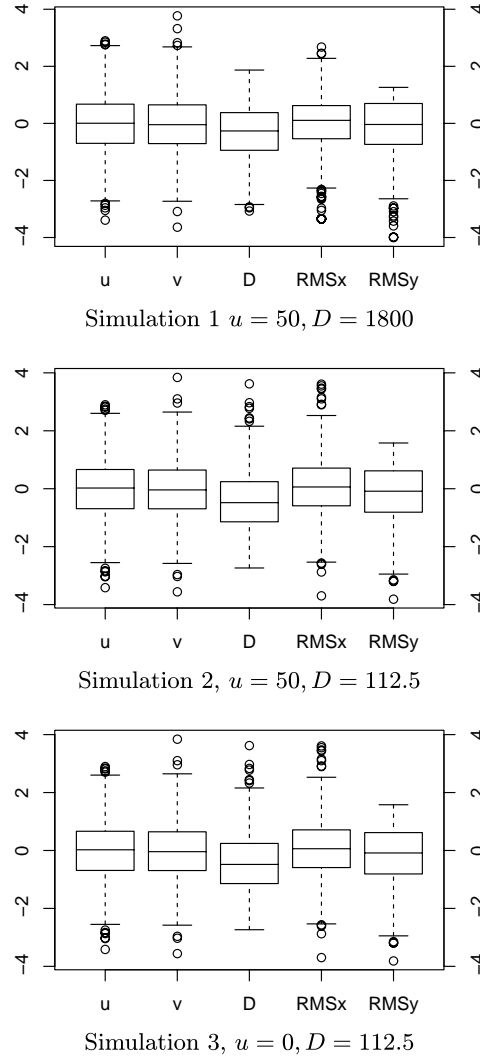
Figure 4 Standardized bias plots. The boxes indicate the interquartile range, i. e., the area encompassing the central 50% of the estimates. The range bars extend outside of the boxes to the most extreme data point which is no more than 1.5 times the interquartile range. The model parameters, $u, v, D, \sigma_x,$ and σ_y are indicated on the abscissa as u, v, D, RMSx and RMSy repectively. The scale on the ordinate in standard deviations.
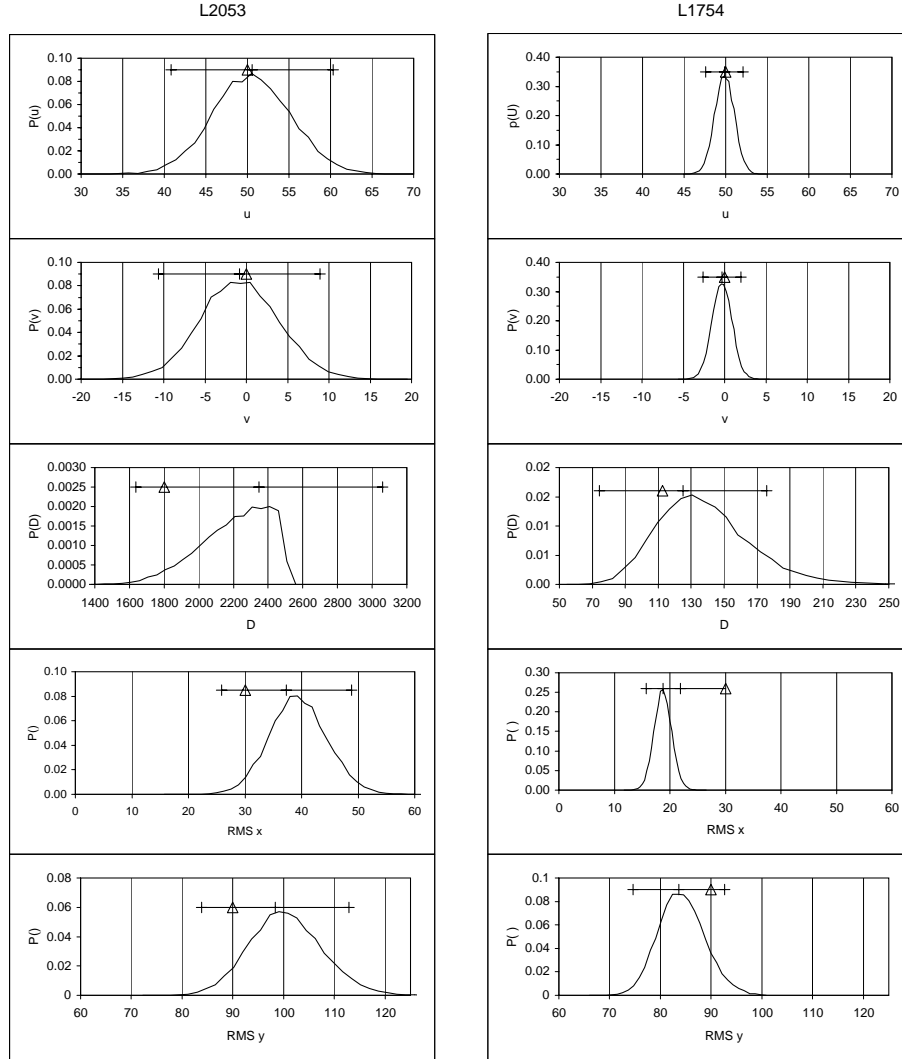
Figure 5   Likelihood profiles for two realizations of simulation. The horizontal bar in each plot represents the point estimate of the parameter $\pm$ two standard deviations. The triangle represents the "true" value of the parameter used in the simulations. L2053 pertains to simulation 1 (high diffusivity). L1754 pertains to simulations 2 and 3 (low diffusivity).

Table 3 Correlation coefficients between parameter estimates. L2053 pertains to simulation 1 (high diffusivity). L1754 pertains to simulations 2 and 3 (low diffusivity).

Parameter	u	v	D	σ_x
			L1754	
v	0			
D	0.0034	0.0032		
σ_x	-0.0031	0.0013	-0.4201	
σ_y	-0.0002	0.0018	-0.0674	0.0283
			L2053	
v	-0.0001			
D	0.0089	-0.0126		
σ_x	-0.0108	0.0069	-0.5499	
σ_y	-0.0028	0.0124	-0.3136	0.1724

The magnitude of the diffusivity in relation to the latitude error appears to influence the bias in the estimates of diffusivity and both components of the geolocation error. The high diffusivity case, L2053, appears to cause overestimation of both diffusivity and the geolocation errors. On the other hand, the low diffusivity case, L1754, produces an accurate estimate of diffusivity, but underestimates the geolocation error.

These relationships are reflected in the correlations coefficient between the parameter estimates, Table 3. The estimates of diffusivity and geolocation error are negatively correlated and the magnitude of the correlation is larger in the high diffusivity case.

6. DISCUSSION

The state space Kalman filter model appears to perform reasonably well as a means to estimate population movement parameters from a track. The model has some difficulty in estimating both diffusivity and geolocation error, however. This difficulty, while potentially troublesome, is not unexpected. In its simplest form, the Kalman filter simply estimates the true position as the weighted mean of the position predicted by the model and the position estimated by the tag with weights inversely proportional to the errors in the two position estimates (Maybeck, 1979). In other words, the Kalman filter is interpreting diffusivity as process error and giving the model prediction less weight in a high-
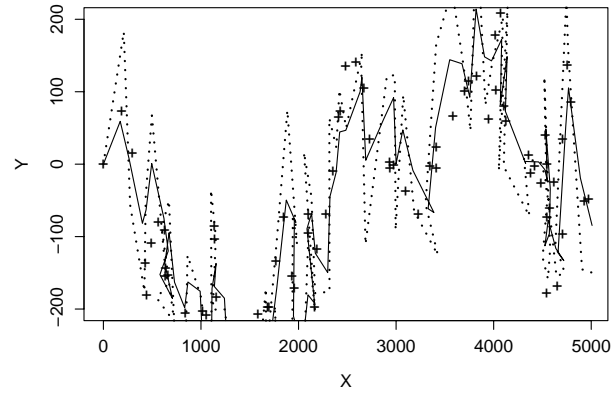
Simulation 1 $u = 50, D = 1800$
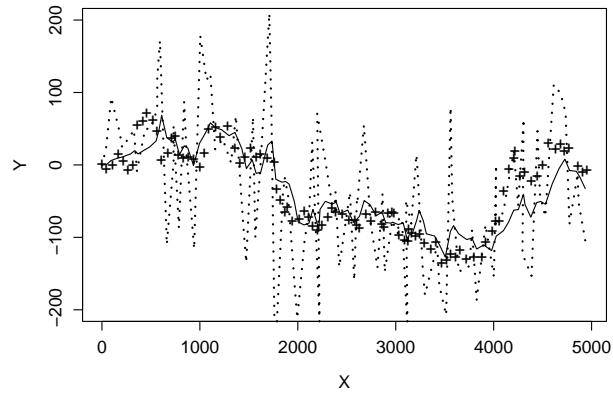
Simulation 2 $u = 50, D = 112.5$
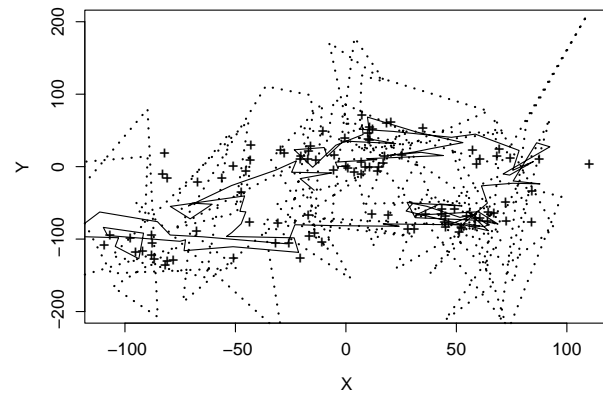
Simulation 3 $u = 0, D = 112.5$

Figure 6 True (+), observed (dotted line) and "corrected" (solid line) position of tags in simulations.

diffusivity track. In spite of this problem, the predicted track is closer to the actual track in all simulations (Figure 6).

The ability of the estimation procedure to correctly evaluate the geolocation error and estimate the true positions of the tagged fish depends on the magnitude of the diffusivity relative to the geolocation error. The geolocation errors used in these simulations are similar to the geolocation errors of the current generation of archival tags operating under optimal conditions and are equivalent to diffusivities of approximately 2000 Nmi^2da^{-1}. During sub-optimal conditions, e.g., around the time of the equinoxes, the geolocation errors would be expected to be much higher (Hill, 2000). Estimates of diffusivity from tagged skipjack in the tropical western Pacific (Sibert et al., 1999) range from 250 to 800 Nmi^2da^{-1}. Similar values were used by Bertignac et al. (1998) for ocean-basin scale of skipjack population dynamics. At the higher range of published tuna diffusivity, Maury et al. (1999) appear to have used values near 9000 Nmi^2da^{-1}in simulations yellowfin tuna populations. These rough comparisons indicate that the diffusivity expected for tunas is probably less than the geolocation errors of archival tags suggesting that the Kalman filter would be applicable to real data from archival tags.

In situations where fish with archival tags and fish with conventional tags are simultaneously at liberty, it is possible to envisage a model which estimates movement and mortality parameters as well as geolocation error in a completely integrated fashion. Such an estimation procedure would link estimation of advection-diffusion and mortality parameters from conventional tags to the estimation of random walk parameters and geolocation errors via a composite likelihood function. Thus, information on the horizontal movements from tracking data would be integrated into large scale population models.

References

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

Feller, W. 1966. An Introduction to Probability Theory and its Applications. Volume II. John Wiley & Sons, N.Y.

Feller, W. 1968. An Introduction to Probability Theory and its Applications. Volume I. (Third Edition). John Wiley & Sons, N.Y. 509 pp.

Hampton, J., K. Bigelow and M. Labelle. 1998. A summary of current information on the biology, fisheries and stock assessment of bigeye tuna (Thunnus obesus) in the Pacific Ocean, with recommendations for data requirements and future research. OFP Tech. Rep. No. 56. Secretariat of the Pacific Community, Noumea.

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

Hill, R. 2000. Geolocation by Light-Level. The next Step: Latitude. Symposium on Tagging and Tracking Marine Fish with Electronic Devices, February 7-11, 2000, Honolulu.

Hinton, M. G. and H. Nakano. 1996. Standardizing catch and effort statistics using physiological, ecological, or behavioral constraints and environmental data, with an application to blue marlin (Makaira nigricans) catch and effort data from Japanese longline fisheries in the Pacific. Bull. Int. Am. Trop. Tuna Comm. 21(4), 171-200.

Lehodey, P., J. Andre, M. Bertignac, J. Hampton, A. Stoens, C. Menkes, L. Mamery and N. Grima. 1998. Predicting skipjack tuna forage distributions using a coupled dynamical bio-geochemical model. Fish. Oceanogr. 7, 317-325.

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

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

Maury, O., D. Gascuel and A. Fonteneau. 2001. Spatial modeling of Atlantic yellowfin tuna population dynamics: a habitat based advection-diffusion-reaction approach with application to the local overfishing study. "Spatial processes and management of fish populations", Anchorage October 27 - 39, 1999. Alaska Sea Grant, in press.

Okubo, A. 1980. Diffusion and Ecological Problems:Mathematical Models. Springer, N.Y. 254 pp.

Sibert, J. R., J. Hampton, D. A. Fournier, and P. J. Bills. 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.

Tsuji, S., T. Itoh, A. Nitta, and S. Kume. 1999. The trans-Pacific migration of a young bluefin tuna, Thunnus thynnus, recorded by an archival tag. Working Paper ISC2/99/15. Interim Scientific Committee for Tuna and Tuna-Like Species in the North Pacific Ocean, January 15 -23, 1999, Honolulu.