Oikos
June 2005; 109(3) : 473 - 484
http://dx.doi.org/10.1111/j.0030-1299.2005.13777.x

The definitive version is available at www.blackwell-synergy.com

A state–space model to derive bluefin tuna movement and habitat from archival tags

F. Royer[a]*, J.-M. Fromentin[a] and P. Gaspar[b]

[a] IFREMER, Centre de Recherche Halieutique Méditerranéen et Tropical, avenue Jean Monnet, BP 171, FR-34203 Sète cedex, France
[b] CLS, Division Océanographie Spatiale, 8-10 rue Hermès, FR-31526 Ramonville St Agne, France.

*: francois.royer@ifremer.f

Abstract: Archival tagging provides a unique way to study the spatial dynamics and habitat of pelagic fish. This technique generates lagrangian data of a particular type in marine ecology: although highly informative about processes at different scales (e.g. horizontal movements versus diving behaviour), such data are impaired by location errors and the lack of combination with actual environmental variability. The present paper introduces a framework for modelling bluefin tuna movement in relation to its habitat, using records of light, depth and temperature from archival tags. Based on data assimilation concepts and methods, we show how an explicit formulation of the observation process and the statistics of external variables (e.g. ambient temperature) can improve precision in geolocation. The proposed method is tested on synthetic data: significant reduction (40 to 50%) in the initial root-mean square error is achieved under different noise scenarios. Assimilating sea surface temperature also allows to perform on-line estimation of a range of observation biases. The performance of the model greatly benefits from the adequate formalisation of different variability sources, and allows potentially to reveal interactions between the fish and its habitat. Using this probabilistic approach, we, however, show that some patterns of interest (e.g. foraging in surface fronts) can hardly be retrieved in a context of large observational and environmental noise.

Keywords: Bluefin tuna, movement, habitat, archival, tag

INTRODUCTION

Northern bluefin tuna (BFT) is a pelagic migratory species distributed in the Atlantic ocean and the Mediterranean sea (Mather, et al. 1995). Its population has undergone a sustained exploitation since the Antiquity, through coastal traps, and recently longlining and purse seining (Fromentin 2003). Like other top predators, its habitat and spatial dynamics are still poorly known, due to the scarcity of direct observations. How large migratory animals make use of the spatio-temporal variability of their physical environment and prey resources is still an open question (Bakun 1996). Knowledge on migration behaviour and long range displacements is not only of particular importance for ecological purposes (see e.g. Ravier and Fromentin 2004), but also for management issues (ICCAT 2003). However, the study of such lagrangian processes remains difficult, both in term of observation and understanding (Nathan, et al. 2003). Advances in this field depend on our ability to link fish behaviour and oceanographic variability over a wide range of temporal and spatial scales. Conventional fishing data (e.g. catches or Catch-Per-Unit-Effort from commercial fleets and research surveys) are either too coarse-grained or collected over insufficiently large areas for this specific purpose. Various dispersal processes as well as demographic and oceanographic stochasticity are embedded in the incomplete space/time snapshots derived from survey or fisheries data (Tyre, et al. 2001). In an analysis of fishery-independent data from aerial surveys and high-resolution remote-sensing, Royer et al. (2004) stressed out the difficulty of inferring ecological processes from static observed patterns or occupancy data.

Archival tags (ATs) and Pop-up Satellite Archival Tags (PSATs) offer a more direct way to study fish behaviour, through the combined records of light (to estimate geolocation), depth and ambient temperature (see Block and Stevens 2001 for a review). While displaying significant drawbacks (e.g. price per unit, premature detachment, satellite transmission failures), this technology has provided valuable insights in the extension and duration of transatlantic migration of BFT, as well as indications about depth and thermal preferences (Block, et al. 2001, De Metrio, et al. 2002, Lutcavage, et al. 1999). Still, these recent results remain largely descriptive and are inferred from a relatively small number of tags. Moreover, individual trajectory is indirectly derived from light-based positions, and is often incomplete or impaired by large observation errors (see e.g. Sibert and Fournier 2001). Challenging tasks are: (1) to reduce uncertainties in locating a tag from light records and other variables, (2) to define a statistical framework taking in account location errors, process errors of movement models and partially observed oceanic variability, to derive useful estimates of global and local (time-varying) movement parameters, (3) to develop protocols to link observed patterns with ecological processes of interest, and (4) to investigate the question of optimally designed tagging strategies. We show here how a general class of Bayesian filters (Evensen 1994, Jonsen, et al. 2003, Morales, et al. 2004) can maximize the output of PSAT data analysis, using non-linear and possibly non-stationary motion models. This approach provides a general framework for jointly exploiting light, environmental and behavioural information from archival tags. First, the theoretical basis of such state-space models is detailed. An example based on simulated PSAT data is then given to assess the performance of the proposed Bayesian filter, using sea surface temperature fields from an Ocean General Circulation Model. We show how such methods can potentially cope with some critical problems induced by archival tagging, conditionally on some specific preliminary work. We seek in particular the inference of high-level information from low-level positional data, to use the track as a random sample of the habitat of the animal.

METHODS

1. State-space modelling

1.1 Generalities

Data assimilation aims at optimally combining information from imperfect models and imperfect measurements (Tarantola and Valette 1984). It provides a sound approach to location estimation when only low-level sensors and inaccurate representations of the environment are available. Data assimilation can be formulated in a Bayesian context, for example using the following parametric state space model:

$$X_{t+1} = f(X_t,\theta_t) + v_t \qquad (1)$$

$$Y_t = g(X_t,\theta_t) + e_t \qquad (2)$$

where X_t is the model state vector including animal location, $f(X_t,\theta_t)$ is a function describing the fish state dynamics, $g(X_t,\theta_t)$ is an observation function, and v_t, e_t are process and measurement errors. θ_t is an unknown parameter vector, potentially dependant on the environment. In the Bayesian framework, solving this problem is equal to estimating a posterior probability density $p(X_t|Y_{1:t})$, where $Y_{1:t}=\{Y_1,\ldots,Y_t\}$ is the historical observation vector. This is done using a two-step algorithm, or filter. A prediction step (eq. 3) is first performed by computing $p(X_t|Y_{1:t-1})$ knowing $p(X_{t-1}|Y_{1:t-1})$ and a transition probability $p(X_t|X_{t-1})$:

$$p(X_t|Y_{1:t-1})=\int p(X_t|X_{t-1})\, p(X_{t-1}|Y_{1:t-1})\,dX_{t-1} \tag{3}$$

The Bayes rule (eq. 4) is then used to update $p(X_t|Y_t)$ with the newly available observation vector $Y_{1:t}$. This is known as the correction step:

$$p(X_t|Y_{1:t})=\frac{p(Y_t|X_t)p(X_t|Y_{1:t-1})}{p(Y_t|Y_{1:t-1})} \tag{4}$$

The density $p(X_t|Y_{1:t})$ then provides an estimate of the hidden state X. In the following part, we give a short presentation of two implementations of the Bayes filter in the linear and general case, as well as their applicability to the geolocation problem in archival tagging.

1.2 The Standard Kalman Filter

The Kalman Filter (or KF, Kalman 1960) provides an analytical solution to the state-estimation of order-1 autoregressive processes, i.e. $p(X_t|Y_{1:t-1})$ of Eq. 3 is estimated sequentially and replaced by $p(X_t|Y_{t-1})$. This filter is the best linear unbiased estimator, and is optimal if the dynamics and the observation process are strictly linear, with normally distributed errors (the formulation of the KF is detailed in Appendix 1). Developed in the signal processing community, the KF has been used recently in fish stock models (Schnute 1994), catch-at-length data analysis (Sullivan 1992) or movement modelling in salmon and bigeye tuna (Newman 1993, Sibert, et al. 2003). This approach has lead to valuable results for the objective analysis of archival tags data, and its main advantage are its ease of implementation in the linear case, and its ability to yield macro-statistics (e.g. the diffusion coefficient in the normal diffusion case).

However, the underlying model of the random walk is based on the diffusion equation which, although flexible and tractable, is difficult to interpret and relies on some unlikely assumptions with little ecological meaning (Turchin 1998). Moreover, as noted by Sibert et al. (2003), the KF suffers from drawbacks in the non-linear and non-stationary case, a common situation in animal ecology when time-varying behaviours are considered (e.g. foraging versus migration). Alternative models are also needed to describe super- or sub-diffusion patterns, as seen around aggregating features or during trans-oceanic migration. Data on ambient temperature need also to be used as additional sources of information (Sibert et al.

2003), thus inducing strong non-gaussianity in the observation densities. Non-gaussianity may also occur in the vicinity of shorelines, since probability distributions must be truncated to exclude position estimates on land. Closed-form analysis of such dynamical systems cannot be achieved using a standard KF. Still, sub-optimal analysis may be performed using the Extended Kalman Filter (EKF) or its recent alternatives, the Ensemble Kalman Filter (EnKF) and the Particle Filter (PF).

1.3 The Monte Carlo Particle Filter

Recently, a more general class of Bayesian filters has been proposed and used successfully to sequentially update the posterior distribution using Monte Carlo methods (Doucet and Godsill 1998). These methods (collectively called Particle Filters or PF) approximate the predicted distribution by a set of Dirac delta functions $\delta(.)$, referred to as an ensemble of particles $X_t^{(i)}$ with corresponding importance weights $w_t^{(i)}$ (Eq. 5).

$$p(X_t|Y_t) \approx 1/N \sum_{1}^{N} w_t^{(i)} \delta(X_t - X_t^{(i)}), \; i=\{1,\ldots,N\} \quad (12)$$

No explicit assumption about the form of $p(X_t|Y_t)$ is made: the PF can be applied to general nonlinear, non-gaussian systems. These non-parametric techniques can be highly efficient in that they allow to focus computing power in regions of high likelihood in the state space. Non-stationary models can also be studied in a straightforward way by extending X with a slowly (compared to the dynamics) time-varying parameter vector. For example, geolocation or temperature measurement bias can be modelled as a weak random walk with fixed or 1/t decaying noise. While the PF is relatively easy to implement, some improvements must be considered, such as prior boosting and resampling strategies to avoid filter degeneracy (Andrieu, et al. 2002).

We provide here an algorithm of the generic Particle Filter which can be applied to any estimation problem (see Doucet and de Freitas, 2001, for more details). For illustration purposes, an iteration of the PF is presented in Fig. 1.

- **Givens:** the initial requirements are a dynamical model (Eqs. 1 and 2) with known state dynamics and observation functions $f(x_t,\theta_t)$ and $g(x_t,\theta_t)$, initial prior densities for the state vector $p(X_0)$ and parameter vector $p(\theta_0)$, and noise densities $p_{v,t}$ and $p_{r,t}$. The filter parameters are the number N of simulated particles and the roughening noise density functions $p_{r,t,x}$ and $p_{r,t,\theta}$ (these are time dependent instrumental noises for smoothing and identification purposes).

- **Initialisation step:** Generate N samples, or particles, from the initial state density $X_0^{(i)} \sim p(X_0)$, with i={1,…,N}. Each particle is given a weight $w_0^{(i)}$:=1/N.
- **Density propagation**: for each particle i={1,…,N}, we sample the state evolution noise $v_t^{X,(i)} \sim p_{v,X}$, the parameter evolution noise $v_t^{\theta,(i)} \sim p_{v,\theta}$, and the associated instrumental noise $r_t^{X,(i)} \sim p_{r,X}$, $r_t^{\theta,(i)} \sim p_{r,\theta}$. The new state and parameter vector are simulated for each particle using Eqs. (13) and (14):

$$X_{t+1|t}^{(i)} = f(X_t^{(i)},\theta^{(i)}) + v_t^{X,(i)} + r_t^{X,(i)} \quad (13)$$

$$\theta_{t+1|t}^{(i)} = \theta_t^{(i)} + v_t^{\theta,(i)} + r_t^{\theta,(i)} \quad (14)$$

- **Assimilation step**: the measurement update is achieved by first updating the weights with a likelihood function for each particle i={1,…,N} :

$$w_t^{(i)} = w_{t-1}^{(i)}\, p(Y_t|X_t^{(i)}) \quad (15)$$

$$w_t^{(i)} = w_{t-1}^{(i)}\, p_{e,t}(Y_t - g(x_t^{(i)},\theta_t^{(i)})) \quad (16)$$

$$w_t^{(i)} = w_t^{(i)}/\Sigma_i\, w_t^{(i)} \quad (17)$$

where Eq. (17) corresponds to weight normalization. The posterior distribution is then obtained by applying Sampling Importance Re-sampling to the weighted particles (Rubin 1987). This is achieved by sampling with replacement N particles from $\{X_t^{(i)}(\theta_t^{(i)},w_t^{(i)})\}$ with probability $w_t^{(i)}$. All samples are then given uniform weights $w_t^{(i)}$=1/N. The distribution of particles at time t now provides an approximation of $p(X_t|Y_t)$ as in Eq. (12). The whole process is then reiterated for time step t + 1. The PF thus constructs a sequential estimate of $X_{1:t}$, as in the KF.

2. Application to fish tracking: a simulated case study

2.1 Introduction

In contrast, Morales, et al. (2004) used a recursive Monte-Carlo technique to fit multiple random walks on Elk movement data, using the WinBUGS package (Spiegelhalter, et al. 1996). The purpose of this software is to ease the use of Markov Chain Monte-Carlo methods in statistical analysis. We chose not to use it here since it is mostly designed to handle parametric density functions which are not available in our case (e.g. ambient temperature), while convergence may become an issue for very large or poorly observed data sets. We therefore implemented a sequential PF using the MATLAB 6.5 software. In our PSAT estimation problem, the state vector is $X_t = \{longitude_t, latitude_t, depth_t, temperature_t\}$ for which observations are available directly (from on-board depth and thermal sensors) or indirectly (through light-base geolocation). Now that the state and observation spaces are specified, a model of the fish dynamics in its environment is needed. We restricted this analysis to a two-dimensional space (the third dimension could be included in another model, as the vertical behaviour of pelagic fishes relies on highly different mechanisms).

2.2 Definition of the process of fish movement

Quantifying the dispersal of organisms is a key step in analysing the distribution and fluctuations of animal populations (Turchin 1991). Thus, a wealth of models has been developed for simulating randomly moving particles on a plane. Purely statistical models such as the classical Random Walk, Fractional Brownian Motion or the Levy Flight (Viswanathan, et al. 2001) can be distinguished from purely mechanistic models. Among them, the speed/angle based model (Bovet and Benhamou 1988, Wu, et al. 2000) is getting increasingly popular, and has found applications in terrestrial ecology (Morales, et al. 2004), avian studies (Thorup, et al. 2000) and marine ecology (Hubbard, et al. 2004). While not fundamentally different from a diffusion process at large time scales, this framework becomes more relevant at local scales, yielding easier to interpret statistics. Speed and turning angle are also differentiable state variables, whereas the drift and diffusion in the classical Random Walk can not be separated locally. By allowing speed/angle correlations to vary between successive time steps, one can generate variable movement patterns according to the migration/foraging paradigm (Ramos-Fernandez, et al. 2003, Atkinson, et al. 2002, Bartumeus, et al. 2002, Mårell, et al. 2002, Viswanathan, et al. 1999, Krakauer and Rodriguez-Girones 1995). Such dual motion model has recently been documented for pelagic fishes, who move in a relatively featureless environment with no long-range visibility (Newlands, et al. 2004). Switching between a "foraging" behaviour and a "migration" behaviour may then be driven by a combination of external variables (e.g. light, prey density) and internal state (e.g. reproduction period). We generated the trajectory of a randomly moving bluefin tuna using the following hierarchical statistical process:

- **Movement mode:** the nature (i.e. forage or migrate) of the movement regime at time t is modelled as an arbitrary binomial process with probability α_t:

$$\text{Mode}_t \sim B(\alpha_t, 1-\alpha_t) \tag{11}$$

- **Speed and azimuth:** for each step t, the individual's speed is sampled from a Weibull density of mean λ_{mode} in (λ_{forage}, $\lambda_{migrate}$) and scale h_{mode} in (h_{forage}, $h_{migrate}$), following the observations of Lutcavage, et al. (2000) on bluefin tuna using ultrasonic telemetry:

$$\text{Speed}_t \sim \text{Weibull}(\lambda_{mode}, h_{mode}) \tag{12}$$

Azimuth deviation between two consecutive steps is sampled from a wrapped Von Mises density with mean 0 and concentration parameter $\kappa_{mode} \in [\kappa_{forage}, \kappa_{migrate}]$, (Thorup, et al. 2000).

$$\text{Angle}_t \sim \text{VonMises}(0, \kappa_{mode}) \tag{13}$$

Intuitively, one can expect migration movements to be more directed and thus, display stronger time correlation, while foraging movements may be more scattered yielding higher residency times. In the stationary case (i.e. constant speed distribution), it is possible to relate such a model to the standard diffusion by computing the expected square displacement using the formulae proposed by Wu, et al. (2000).

2.3 Synthetic data generation and scenarios testing

To illustrate the proposed data assimilation method for analysing archival tags, we assimilated ambient temperature in a two-dimensional case study.

- **Simulated track** : Movement was generated over 240 days between the Gibraltar strait and the mid-Atlantic, with a random alternation of directional and adirectional motion regimes (Fig. 2). Time step duration was set to one hour, consistently with the recording resolution in pop-up archival tagging. A geolocation process was then simulated by subsampling the trajectory at a 24-hour period and adding gaussian noise (error sources relate to the time resolution of the recordings, the vertical behaviour of the fish, cloudiness, the proximity of equinoxes, and the optical characteristics of the upper layers (Musyl, et al. 2001 and Welch and Eveson 1999)). Biases were also added to account for possible biofouling or alteration of the light sensors thus inducing drifts in the geolocation process. Finally, records of ambient temperature were generated using the sea surface temperature of a weekly output of Mercator PSY-2, an eddy-resolving resolution Ocean General Circulation Model (Bahurel, et al. 2002). The temperature fields were smoothed with low-pass filters of various bandwidths, to simulate typical uncertainties and coarse resolution in oceanographic data.

- **Scenarios** : The outputs of the Particle Filter assimilating SST were compared to the standard Kalman filter (Sibert and Nielsen 2004) under different noise and bias cases. Issues in latitude estimation were especially investigated by adding gaussian errors ranging from 1 to 3 degrees, while error in estimated longitude were kept low (from 0.5 to 1 degree), a common feature observed in PSATs. Biases in light level sensing/latitude geolocation were simulated as linear drifts (e.g. to account for sensor biofouling) or abrupt shifts (e.g. to account for deterioration) from 1 to 5 degrees in latitude. The SST fields used for temperature assimilation were smoothed to obtain Root-Mean Square Errors (RMSEs) ranging from 0 to 0.5 Celsius degrees, which is the magnitude of the average error in most AVHRR-derived SST products. As the true state of the system was always known, the performance of each method was assessed by computing the RMSE between the estimated state (i.e. the average of the estimated density) and the true (hidden) state time-series.

RESULTS

1. Comparison of the KF and the PF

As a verification, we first applied the KF and the PF to state vectors of comparable dimension, i.e. without SST assimilation. Similar random walk models were used for the propagation dynamics, i.e. an advection-diffusion scheme for the KF and its discrete ensemble counterpart for the PF, while all observation errors were held gaussian. Both filters yielded similar results in terms of hidden track estimate and RMSE reduction (results not shown).

We then applied the PF and the KF to unbiased series of observations with different errors in the longitude and latitude components, and given SST statistics for the PF. Two examples with low and high geolocation errors are provided as an illustration in Fig. 3. The observation error was set to 0.5 and 1 degree in longitude and latitude for the low deviation scheme, and 1 and 3 for the large deviation scheme. SST error was set to a very low value

(0.01°C), which stands for almost perfect *a posteriori* knowledge of the SST field (errors in the forcing SST field are investigated further). Both the KF and PF solutions allowed for a significant decrease in the RMSE of the estimated trajectory compared to observed positions. For weak observation errors, the decrease in the estimate RMSE reached 38% and 50% for the KF and PF, respectively. While the PF yielded a better solution, there were little differences between the KF and the PF estimated tracks (Fig. 3a). The standard Kalman Filter was also outperformed by the Particle Filter for greater observation errors: the observation RMSE was decreased by 42% and 58%, respectively. On average, both filters performed well in this case but qualitatively, the PF estimate still followed accurately the true track, while the KF yielded a smoother estimate (i.e. the KF solution missed a number of local features in the original track –see Fig. 3b). For both cases, the PF solution for the temperature remained almost identical to the true record, as shown in the inserts of Fig. 3. Such result is logical, since a very high confidence was given to temperature data at the assimilation step. Generally, the PF was applied here to a three-dimensional system, thus benefiting from a greater amount of data for its task.

2. Assimilating SST

We then investigated to which extent the observation errors were reduced when assimilating SST of varying accuracy. The performance of the PF was expressed as a function of the RMSE of the observed positions as well as the RMSE of the smoothed SST input. About 1000 trials were run with observation errors ranging from 0.5 to 3 degrees and SST error from 0 to 0.5 °C. The propagation of input uncertainty was represented as a response surface (Fig. 4). The observed positions RMSE ranged from 0.8 to 3 degrees, while the PF solution (i.e. estimated track) displayed errors ranging from 0.6 to 0.95 degrees, so that reduction in RMSE attained 25% to 68%. Error on the SST field (in the range studied) had globally little effect on the RMSE of the PF solution: for a given geolocation RMSE, an SST error ranging from 0 to 0.5°C yielded a variation of about 0.1 degree in the final geolocation error. For low geolocation errors (i.e. less than 2 degrees), an increasing SST error increased only slightly the RMSE of the PF solution. For geolocation errors greater than 2 degrees, the opposite situation occurred: low SST RMSE (i.e. less than 0.3 °C) lead to higher error in the PF solution than for high SST RMSE (i.e. more than 0.3 °C). In other words, high confidence in the knowledge of the temperature field degraded the performance of the filter when combined with noisy geolocations, while less accurate (i.e. smoother) fields lead to better results in this case. As a conclusion, SST error had relatively little effect on the PF estimated track, whereas geolocation error (in longitude, latitude) had the greatest influence. This response surface was obtained with informative prior distributions on both geolocation error and individual motion: misleading priors naturally decreased the performance of the PF.

3. Estimating geolocation biases

The PF clearly outperformed the KF when supplied with biased geolocations: three examples are given assuming a fixed latitude bias of 2 degrees (Fig 5a), a linear increase from 0 to 5 degrees (Fig 5b), and an abrupt shift at mid-track from 0 to 5 degrees (Fig 5c). These bias schemes were supposed to be unpredictable, i.e. relying on unknown causes such as biofouling or sensor alteration. The KF performed well for a fixed bias, since the tagging location and jettison point provide useful information for its correction (Fig. 5a). However this information was not well propagated in time for a linear drift or an abrupt shift (Fig. 5b, 4c), yielding to globally poor estimates of the true tracks. When supplied with unbiased temperature data, the PF provided satisfying solutions for biases in observed latitude. For example, a constant bias was identified with 1/t time decaying noise, providing a converging solution (Fig 5a). Time-varying biases were identified with a constant noise, yielding rougher solutions (Figs 5b and 5c), to allow for drift identification and shifts detection. Assimilating external temperature apparently gave a decisive advantage over the standard KF for bias estimation, especially at mid-track when tagging location or jettison point were distant in time.

4. Inferring movement patterns

Using three observable components of the system (i.e. longitude, latitude and ambient temperature), the PF performed well in correcting the observation process, by 1) increasing the signal-to-noise ratio, and 2) estimating a non-observed variable such as the latitude bias. We then assessed its performance in retrieving hidden variables of the dynamical model itself: Fig. 6 displays a typical PF solution for a true track observed with large latitude errors (> 2 degrees), with the estimated and true movement regime time series (insert in Fig. 6). The alternation of directional and adirectional movement was globally not well retrieved (the second part of the true track being totally directional). Still, this output illustrates a number of false detections, e.g. segments of the fitted trajectory are identified as originating from directional motion whereas the corresponding true segment was adirectional. Such result can be observed in segments of high sinuosity, where a large number of adirectional detections occurred. The identification of such hidden variable was highly sensitive to initial priors and ensemble size. Percentages of good identifications reached 70% for small geolocation errors, while it was not significantly different from 50% for higher geolocation errors (e.g. more than 2 degrees). Thus, for high observation errors, the PF did not yield better results than a blind guess in the retrieving the hidden variables of the dynamical system.

DISCUSSION

The proposed Monte Carlo filter performed well in filtering the track of an individual in a 2-D geographical space (i.e. longitude and latitude). When applied to a linear random walk and gaussian errors, the Particle Filter with Sampling-Importance-Resampling step was comparable to the Kalman Filter. This validated the approximation scheme and the representation of probability distributions by an ensemble of particles. The PF was then extended to a higher observation space (i.e. longitude, latitude and temperature), and a more complex random walk model with regime shifts. This induced strong non-gaussianity in the state distributions, particularly in the latitudinal dimension (see examples in Fig. 7). Multimodal and highly skewed distributions were generated by sea surface temperature, a feature that could be handled by the non-parametric nature of the PF.

Including SST as a supplementary variable in the space-time analysis allowed a higher RMSE reduction than in the standard Kalman filter. For high observation errors, the KF behaved like a smoother since geolocation data contributes less to the variability of the final solution. Temperature assimilation prevented excessive relaxing in the PF estimate, thus giving insight in the true variability of the track. Obviously, the PF handled a greater amount of information, since it was applied to a space of higher dimension (i.e. longitude, latitude and temperature). The PF’s ability to handle a variety of parametric/non-parametric state and parameters distributions is one of its key features. However, its performance is conditioned by two major issues. First, external forcing variables such as the temperature field need to be spatially informative, and second, the statistics of the observation errors need to be well known for the PF to perform correctly. We illustrated this by using SST fields smoothed with gaussian filters of varying bandwidths. This allowed to preserve their spatial pattern: thus, the temperature RMSE had little effect on average on the RMSE of the PF solution. The

Pathfinder AVHRR-derived SST, at its current status, presents a bias/error scheme of 0.00±0.24°C, which is more than adequate for such analysis (Keogh, et al. 1999). Interestingly, the combination of high observation error and high precision in temperature generates a high error in the PF estimate. This is interpreted as a greater spatial indetermination of the system since sea surface temperature may display pseudo-periodic properties, thus generating multiple choices. As a general result, geolocation uncertainty may cover areas where several water masses display a suitable temperature distribution, thus generating multimodal densities (see Fig. 7). The probability density functions of these state estimates were retrieved by applying gaussian kernels to the updated ensembles, thus allowing non-parametric representations of the probable dispersion in latitude. Issues about the SST spatial structure could be explored *a priori* by comparing geolocation errors (in both latitude and longitude) with empirical variograms or spatial autocorrelation of the SST field.

On-line estimation of hidden variables was also performed, both in the observation and state space. Retrieving time-varying bias in latitude, for example, was achieved by allowing the model parameters to change with time as a random walk. A fixed bias was well identified by using a 1/t time decaying noise, thus providing converging estimates. Shifts in the bias were identified using randomly walking parameters (i.e. with noise constant in time): this allowed to detect changes in bias, at the cost of a rougher estimate. Regime shifts identification therefore relies on the specification of a suitable noise model in hidden variables. Such estimation is achieved using information on the tag's initial and final positions and the temperature field (it is therefore critical that prior SST fields are free of strong bias for this task). Their accuracy must be assessed as a preliminary work, since it is difficult to estimate both geolocation errors and modelled SST bias at the same time in state-space models. Regime shifts in the dynamic process itself (i.e., migration vs foraging behaviour) were also estimated, yielding less satisfying results: a number of false detections occurred, in relation to both model specification (e.g. autocorrelation of regimes in time) and geolocation errors. This was related to a greater indetermination of the system, i.e. several combinations of parameters have equal probabilities to yield the observed pattern in the data. In this case, the PF solution was not significantly better than a blind guess of the movement regime (i.e. equal probability of choosing between a "directional" and an "adirectional" step). Moreover, the PF appeared to favour the "adirectional" movement, since this least restrictive regime may have yielded higher likelihoods. This feature can be directly related to inverse diffusion, a typically ill-defined problem in which a given state/observation can originate from multiple sets of parameters or initial states. Overcoming this problem depends on reducing the degree of freedom of the system, i.e. through a better formulation of system noise and hidden processes. Supplemental knowledge could be added, as derived from direct observation, such as swimming speed distributions in feeding grounds, frontal zones, or during migration periods. This would also allow to include testable hypothesis in the analysis of archival tag data, by linking the behaviour of the individual to its immediate surroundings. The PF offers a flexible tool for estimating the trajectory of individuals with an assimilation of data on their physical environment, but it is not without any drawback, since its great flexibility can be a handicap for process identification (i.e. the system can quickly become under-determined when supplied with insufficient data). Such drawback was illustrated by the poor estimation of hidden processes variables (i.e. regime shifts in movement statistics): the combined estimation of states and parameters is not easily performed in a single step in a Bayesian framework. Statistical analysis of the track may be performed as a subsequent step, e.g. the most probable trajectory can then be summarized through standard lagrangian statistics, or used in posterior studies for deriving environmental preferences, or assessing interactions with various oceanic features (e.g. fronts and eddies).

Linking fish behaviour to its habitat relies on some optimal knowledge of environmental heterogeneity: we therefore need to describe oceanographic variability over a sufficient range of scales. Interfaces between water masses and transition/frontal zones can be of particular interest, since ecological processes of great importance occur at these locations (Bakun 1996). The proposed lagrangian method would allow to locate a fish in relation to its habitat over periods of days to months, thus providing insights on how top predators with high metabolism can interact with the high patchiness of their environment. Species such as bluefin tuna can be seen both as mobile sampler and integrator of oceanic variability, having developed specific strategies to preserve homeostasis. Thermoregulation, fat storage are such well-known physiological mechanisms. Specific foraging strategies are also believed to maximize expected food intake, through opportunistic feeding or Brownian search. Still, little quantitative information is available on how pelagic predators find food in space: integrated studies linking behavioural knowledge, archival data and oceanographic variability may provide interesting clues about how large fish perceive the ocean landscape. The random walk based on speed/angle variations may allow some insight on how pelagic animals orientate in the pelagic domain. Accurate models of tuna behaviour are especially needed in such analyses: such flexible motion models, although conceptually satisfying and versatile, may not be enough informative, thus forbidding data enrichment and yielding data-driven estimates. Data assimilation theory was elaborated in fields such as applied physics, where processes are well known and described by dynamical equations. Unfortunately, this is not the case in ecology, where the actual dynamics of individuals are most often best described by statistical or empirical models.

CONCLUSION

We presented in this paper a general PF algorithm that can be applied to parameter-state estimation in animal behaviour studies. Using an ensemble representation of state distributions and sampling-importance-resampling for prior update, we were able to filter the observed trajectory of an instrumented fish. The flexible, non-parametric nature of the PF also permitted the assimilation of external data of general type (here sea surface temperature), thus allowing integrated bio-physical oceanographic studies. We derived a range of observation scenarios from various sets of simulated errors: this pointed out the need for accurate knowledge of the underlying processes and their variability. Gathering data and formalizing hidden processes is an important preliminary step in Bayesian analysis. Only then may the output of such inverse modelling studies be maximized: it was shown in particular that process identification is conditioned by the reduction of observation errors and a clear formulation of the modelled dynamics. This approach opens the gate to a variety of integrated studies at the crossroads of marine behavioural ecology and physical oceanography.

ACKNOWLEDGEMENTS

We thank Olivier Maury who provided helpful comments and advice on a earlier version of this manuscript. This research was made possible by a PhD grant funded by CLS and IFREMER for FR.

REFERENCES

Andrieu, C., Davy, M. and Doucet, A. 2002. Efficient Particle Filtering for Jump Markov Systems. - IEEE transactions on Signal Processing 51: 1762-1769.

Atkinson, R. P. D., Rhodes, C. J., Macdonald, D. W. and Anderson, R. M. 2002. Scale-free dynamics in the movement patterns of jackals. - Oikos 98: 134-140.

Bahurel, P., De Mey, P., Le Provost, C. and Le Traon, P. Y. 2002. A GODAE prototype system with applications: Example of the Mercator system. "En route to GODAE" 2002 Proceedings - CNES (ed.) France

Bakun, A. 1996. Patterns in the Ocean: Ocean Processes and Marine Population Dynamics. - University of California Sea Grant, San Diego, California, USA, in cooperation with Centro de Investigaciones Biológicas de Noroeste, La Paz, Baja California Sur, Mexico.

Bartumeus, F., Catalan, J., Fulco, U. L., Lyra, M. L. and Viswanathan, G. M. 2002. Optimizing the encounter rate in biological interactions: Lévy versus Brownian strategies. - Physical Review Letters 88.

Block, B. A., H., D., Blackwell, S. B., Williams, T. D., Prince, E. D., Farwell, C. J., Boustany, A., Teo, S. L. H., Seitz, A., Andreas, W. and Fudge, D. 2001. Migratory Movements, Depth Preferences, and Thermal Biology of Atlantic Bluefin Tuna. - Science 293: 1310-1314.

Block, B. A. and Stevens, E. D. 2001. Tuna. Physiology, ecology, and evolution. - Academic Press.

Bovet, P. and Benhamou, S. 1988. Spatial analysis of animal movements using a correlated random walk model. - J. Theor. Biol. 131: 419-433.

De Metrio, G., Arnold, G. P., Block, B. A., De la Serna, J. M., Deflorio, M., Cataldo, M., Yannopoulos, C., Megalofonou, P., Beemer, S., Farwell, C. and Seitz, A. 2002. Behaviour of post-spawning atlantic bluefin tuna tagged with pop-up satellite tags in the mediterranean and eastern atlantic. - Col. Vol. Sci. Pap. ICCAT 54: 415-424.

Doucet, A. and Godsill, S. 1998. On Sequential Monte Carlo Sampling Methods for Bayesian Filtering. - Cambridge University.

Evensen, G. 1994. Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods for forecast error statistics. - J. Geophys. Res. 99: 10143-10162.

Fromentin, J.-M. 2003. The East Atlantic and Mediterranean Bluefin tuna management: uncertainties and alternatives. - Scientia Marina 67: 51-62.

Hubbard, S., Babak, P., Sigurdsson, S. T. and Magnusson, K. G. 2004. A model of the formation of fish schools and migrations of fish. - Ecol. Modell. In press.

ICCAT. 2003. Report of the 2002 Atlantic bluefin tuna stock assessment session. - Col. Vol. Sci. Pap. ICCAT 55: 710-937.

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. 1960. A New Approach to Linear Filtering and Prediction Problems. - Transactions of the ASME--Journal of Basic Engineering 82: 35-45.

Keogh, S. J., Robinson, I. S., Donlon, C. J. and Nightingale, T. J. 1999. The accuracy of AVHRR SST determined using shipborne radiometers. - International Journal of Remote Sensing 20: 2871 - 2876.

Krakauer, D. C. and Rodriguez-Girones, M. A. 1995. Searching and learning in a random environment. - J. Theor. Biol. 177: 417-419.

Lutcavage, M., Brill, R., Skomal, G. B., Bradford, C. 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.

Lutcavage, M., Brill, R., Skomal, G. B., Chase, B. C., Goldstein, J. and Tutein, J. 2000. Tracking adult North Atlantic bluefin tuna (*Thunnus Thynnus*) in the northwestern Atlantic using ultrasonic telemetry. - Mar. Biol. 137: 347-358.

Mårell, A., Ball, J. P. and Hofgaard, A. 2002. Foraging and movement paths of female reindeer: insights from fractal analysis, correlated random walks, and Lévy flights. - Can. J. Zool. 80: 854-865.

Mather, F. J., Mason, J. M. and Jones, A. C. 1995. Historical document: life history and fisheries of Atlantic bluefin tuna. - NMFS-SEFSC, p. 165 pp.

Morales, J. M., Haydon, D. T., Frair, J., Holsinger, K. E. and Fryxell, J. M. 2004. Extracting more out of relocation data: building movement models as mixtures of random walks. - Ecology in press.

Musyl, M. K., Brill, R., Curran, D. S., Gunn, J. S., Hartog, J. R., Hill, R. D., Welch, D. W., Eveson, J. P., Boggs, C. H. and Brainard, R. E. 2001. Ability of archival tags to provide estimates of geographical position based on light intensity. - Reviews in Fish Biology and Fisheries In press.

Newlands, N. K., Lutcavage, M. and Pitcher, T. J. 2004. Analysis of foraging movements of bluefin tuna (*Thunnus thynnus*). Individuals switch between two modes of search behavior. - Population Ecology in press.

Newman, K. B. 1993. State-Space Modeling of salmon migration and Monte Carlo alternatives to the Kalman filter. - University of Washington.

Ramos-Fernandez, G., Mateos, J. L., Miramontes, O., Cocho, G., Larralde, H. and Ayala-Orozco. 2003. Lévy walk patterns in the foraging movements of spider monkeys (*Ateles geoffroyi*).

Ravier, C. and Fromentin, J.-M. 2004. Are the long term fluctuations in Atlantic bluefin tuna population related to environmental changes? - Fisheries Oceanography 13: 145-160.

Royer, F., Fromentin, J.-M. and Gaspar, P. 2004. Association between bluefin tuna schools and oceanic features in the western Mediterranean. - Mar. Ecol. Prog. Ser. 269: 249-263.

Rubin, D. B. 1987. A noniterative sampling/importance resampling alternative to the data augmentation algorithm for creating a few imputations when fractions of missing information are modest: the SIR algorithm. Discussion of "The calculation of posterior distributions by data augmentation" by Tanner and Wong. - Journal of the American Statistical Association 82: 543-546.

Schnute, J. 1994. A general framework for developing sequential fisheries models. - Can. J. Fish. Aquat. Sci. 51: 1676-1688.

Sibert, J. R. and Fournier, D. A. 2001. Possible models for combining tracking data with conventional tagging data. Symposium on tagging and tracking marine fish with electronic devices Proceedings - Honolulu

Sibert, J. R., Musyl, M. K. and Brill, R. 2003. Horizontal movements of bigeye tuna (Thunnus obesus) near Hawaii determined by Kalman filter analysis from archival tagging data. - Fisheries Oceanography 12: 141-151.

Sibert, J. R. and Nielsen, A. 2004. The KFtrack package.

Spiegelhalter, D. J., Thomas, A., Best, N. and Gilks, W. R. 1996. BUGS 0.5, Bayesian inference using Gibbs sampling. - Medical Research Council Biostatistics Unit, Institute of Public Health.

Sullivan, P. 1992. A Kalman filter approach to catch-at-length analysis. - Biometrics 48: 237-257.

Tarantola, A. and Valette, B. 1984. Inverse problems - Quest for information. - J. Geophys. 50: 159-170.

Thorup, K., Rabøl, J. and Madsen, J. J. 2000. Can clock-and-compass explain the distribution of ringing recoveries of pied flycatchers? - Animal Behaviour 60: 3-8.

Turchin, P. 1991. Translating foraging movements in heterogeneous environments into spatial distribution of foragers. - Ecology 72: 1253-1266.

Turchin, P. 1998. Quantitative analysis of movement. - Sinauer.

Tyre, A. J., Possingham, H. P. and Lindenmayer, D. B. 2001. Inferring process from pattern: can territory occupancy provide information about life history parameters? - Ecological Applications 11: 1722-1737.

Viswanathan, G. M., Afanasyev, V., Buldyrev, S., Havlin, S., da Luz, M. G. E., Raposo, E. P. and Stanley, H. E. 2001. Lévy flights search patterns of biological organisms. - Physica A. 295: 85-88.

Viswanathan, G. M., Buldyrev, S., Havlin, S., da Luz, M. G. E., Raposo, E. P. and Eugene Stanley, H. 1999. Optimizing the success of random searches. - Nature 401: 911-914.

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

Wu, H., Li, B.-L., Springer, T. A. and Neill, W. H. 2000. Modelling animal movement as a persistent random walk in two dimensions: expected magnitude of net displacement. - Ecol. Modell. 132: 115-124.

Figures

Figure 1. Iteration of a Particle Filter between t-1 and t for N=10 particles.

Figure 2. Synthetic trajectory of a freely-swimming bluefin tuna. The shift between motion regimes (i.e. directional/adirectional) is governed by an arbitrary binomial process. This random walk has been simulated over 240 days with a 1-hour time step. Insert shows the corresponding along-track temperature sampled from a weekly output of the MERCATOR circulation model.

Figure 3. Comparison of the standard Kalman Filter estimate (blue line) and the Particle Filter estimate (dark line) for two unbiased observation processes with error of: a) 0.5 degree in longitude and 1 degree in latitude, b) 1 degree in longitude and 3 degrees in latitude. Actual (hidden) track is in red; observations are materialized by grey crosses.

Figure 4. Response surface of the Root Mean Square Error (RMSE) of the Particle Filter output, as a function of the initial geolocations RMSE and the temperature field RMSE.

Figure 5. Three scenarios of bias estimation in observed latitude using the Particle Filter: a) with a constant bias of 2 degrees, b) with a linearly increasing bias from 0 to 5 degrees, and c) with an abrupt change from 0 to 5 degrees at mid-track. The ensemble estimate of the bias (black thick line) is given along with ensemble 95% confidence intervals (grey lines). The dashed line stands for the true (hidden) bias. The maps on the right show the corresponding biased observations (+), with the KF most probable track (grey line) and the PF estimated track (dark line).

Figure 6. Regime shift detection along the simulated track, observed with errors of 0.5 degree in longitude and 1 degree in latitude. Black line is the PF estimated track, dashed line is the true (hidden) track. Observed geolocations are symbolized by crosses. Black dots point out segments estimated as “adirectional” by the PF. Insert shows the time series of the true (hidden) sequence of regimes and the PF estimated sequence, with black lines representing adirectional movement.

Figure 7. Latitudinal distribution of the PF-estimated daily geolocations (grey curves). Black dots symbolize the daily ensemble means.

Appendix 1. Formulation of the Standard Kalman Filter

Given a linear system yielding observations Y of a dynamical state X:

$$X_t = AX_{t-1} + BU_{t-1} + v_{t-1} \tag{5}$$

$$Y_t = HX_t + e_t \tag{6}$$

Where U_{t-1} is the drift, $v \sim N(0,Q)$ is a process noise of variance Q, and $e \sim N(0,R)$ is an observation noise of variance R. As the errors are assumed gaussian, the density of the state vector is fully described by its two first moments, i.e. its mean state X and covariance matrix P. The filter itself consists first of a time update of both X and P (prediction step):

$$X_{t|t-1} = AX_{t-1} + BU_{t-1} \tag{7}$$

$$P_{t|t-1} = AP_{t-1}A^T + Q \tag{8}$$

The measurement update (or assimilation step) is then performed by computing the Kalman gain K_t and updating both X and P, using the equations derived by (Kalman 1960):

$$K_t = P_tH^T(HP_tH^T + R)^{-1} \tag{9}$$

$$X_t = X_{t|t-1} + K(Y_t - HX_{t|t-1}) \tag{10}$$

$$P_t = (I - K_tH)P_{t|t-1} \tag{11}$$

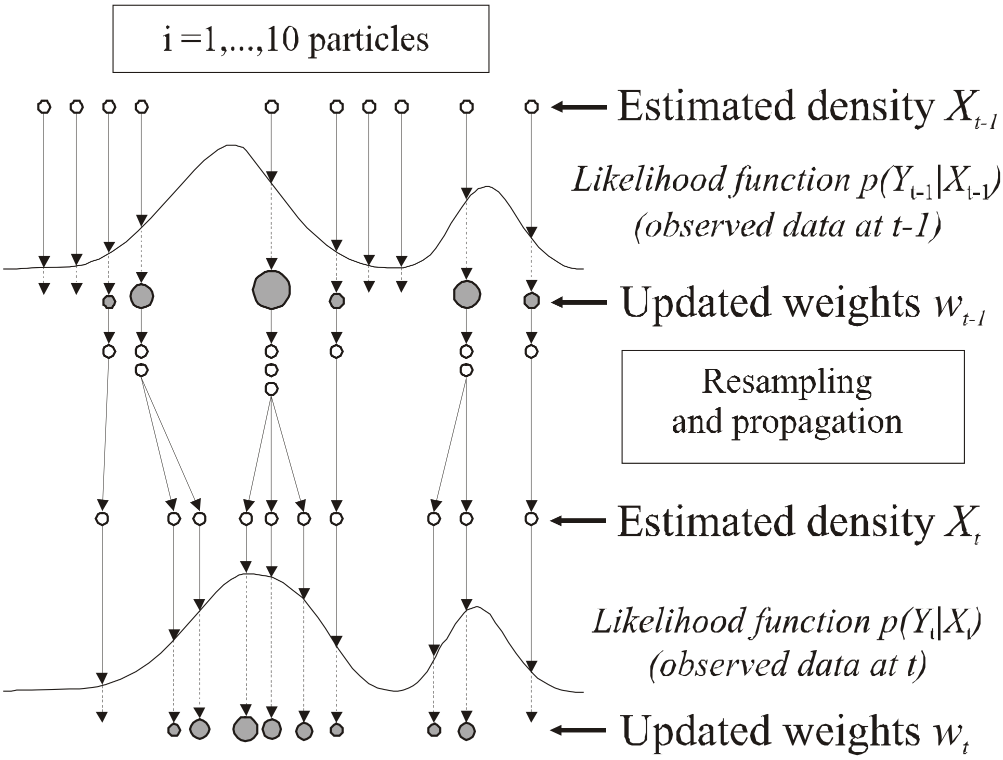

i =1,...,10 particles
Estimated density X_{t-1}
Likelihood function $p(Y_{t-1}|X_{t-1})$
(observed data at t-1)
Updated weights w_{t-1}
Resampling
and propagation
Estimated density X_t
Likelihood function $p(Y_t|X_t)$
(observed data at t)
Updated weights w_t

Figure 1

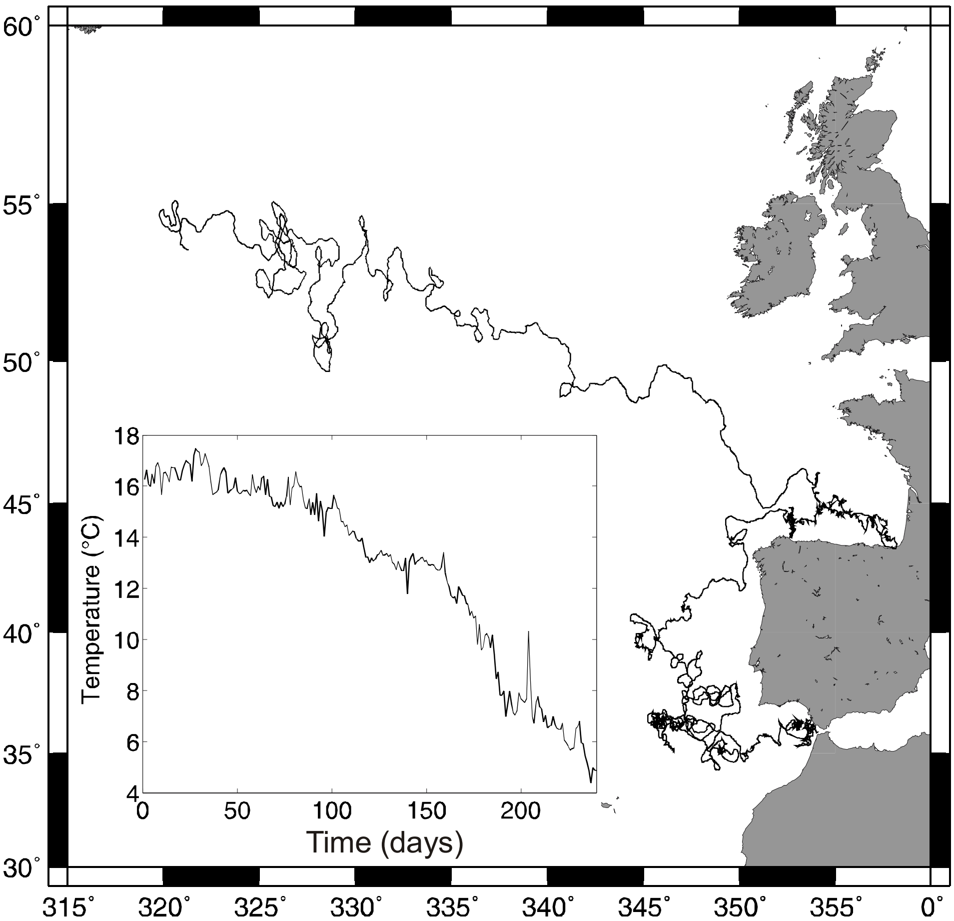

60°
55°
50°
45°
40°
35°
30°
315°
320°
325°
330°
335°
340°
345°
350°
355°
0°
Temperature (°C)
Time (days)
18
16
14
12
10
8
6
4
0
50
100
150
200

Figure 2

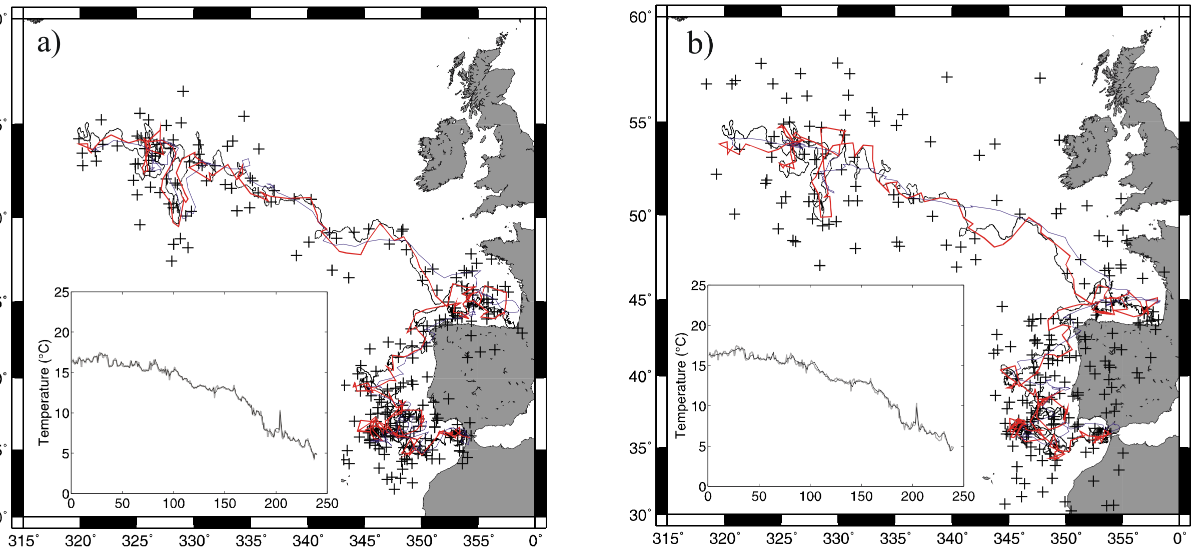
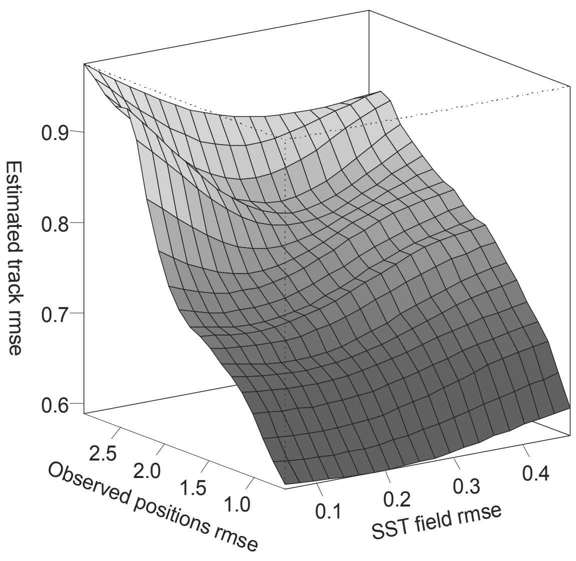

a)
b)
60°
55°
50°
45°
40°
35°
30°
315°
320°
325°
330°
335°
340°
345°
350°
355°
0°
Temperature (°C)
25
20
15
10
5
0
0
50
100
150
200
250

Figure 3



Estimated track rmse
0.9
0.8
0.7
0.6
2.5
2.0
1.5
1.0
Observed positions rmse
0.1
0.2
0.3
0.4
SST field rmse

Figure 4

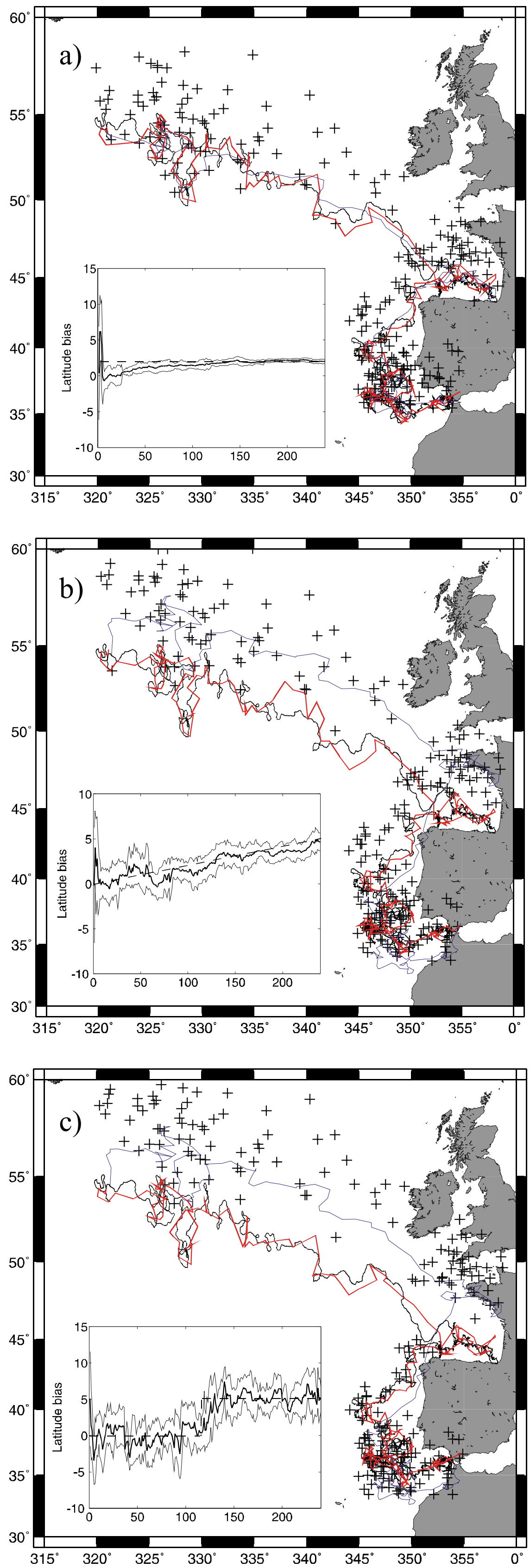

a)
b)
c)
Latitude bias
60˚
55˚
50˚
45˚
40˚
35˚
30˚
315˚
320˚
325˚
330˚
335˚
340˚
345˚
350˚
355˚
0˚

Figure 5

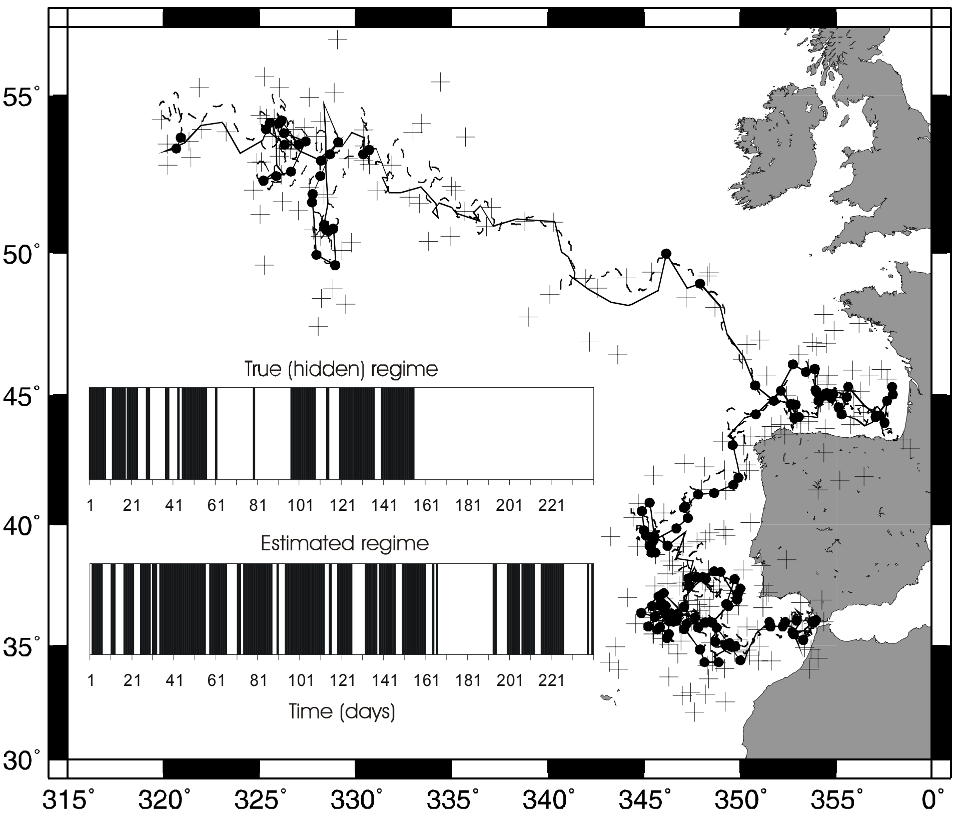

True (hidden) regime
Estimated regime
Time (days)
1 21 41 61 81 101 121 141 161 181 201 221
55˚
50˚
45˚
40˚
35˚
30˚
315˚ 320˚ 325˚ 330˚ 335˚ 340˚ 345˚ 350˚ 355˚ 0˚

Figure 6

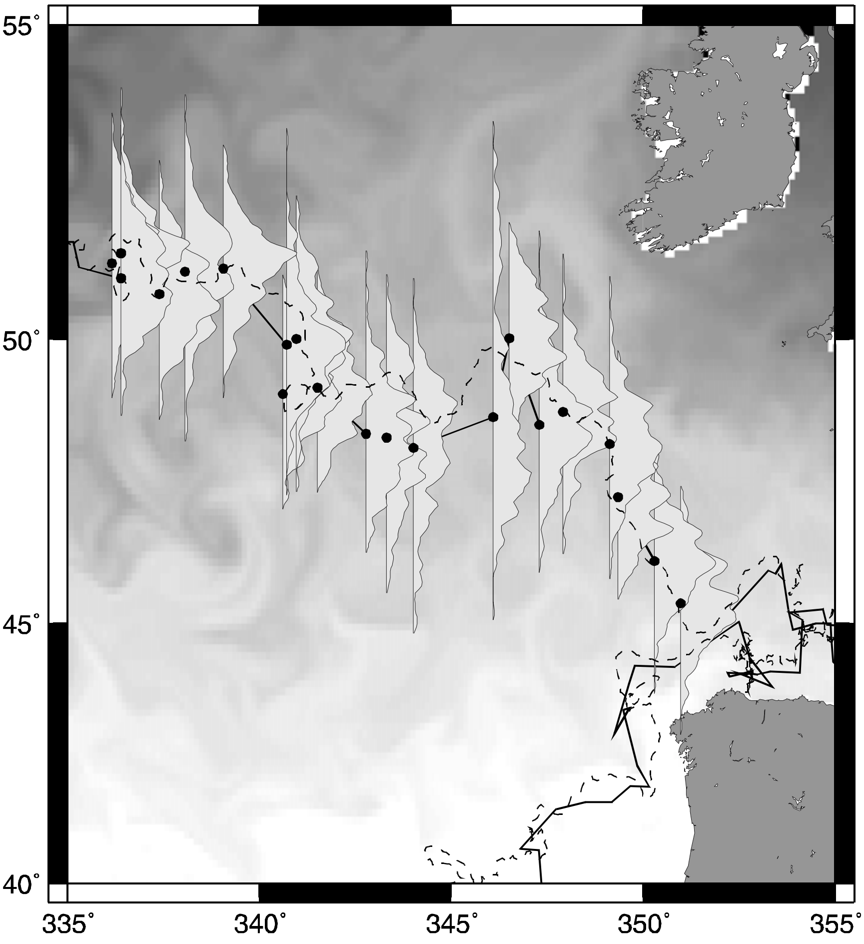


Figure 7