

# Using the particle filter to geolocate Atlantic cod (*Gadus morhua*) in the Baltic Sea, with special emphasis on determining uncertainty

K.H. Andersen, A. Nielsen, U.H. Thygesen, H.-H. Hinrichsen, and S. Neuenfeldt

**Abstract:** The use of archival tags on fish gives information of individual behaviour with an unprecedented high resolution in time. A central problem in the analysis of data from retrieved tags is the geolocation, namely the inference of movements of the fish by comparing the data from the tags with environmental observations like temperature, tide, day length, etc. The result is usually represented as a track; however, the spatial and temporal variability in the precision is often substantial. In this article, the particle filter is applied to geolocate Atlantic cod (*Gadus morhua*) in the Baltic Sea, leading to a representation of the results as probability distributions for each time step, thus giving an explicit representation of uncertainty. Furthermore, the method is used to estimate the magnitude of the error in the measurements by the tags and the swimming velocity of the fish. The average swimming velocity during a day was estimated to be around  $0.20 \text{ m}\cdot\text{s}^{-1}$  for fish of  $\sim 60 \text{ cm}$  length. The method is general and the presentation is formulated to facilitate implementation for different systems where other quantities are observed.

**Résumé :** L'utilisation d'étiquettes enregistreuses donne des renseignements sur les comportements individuels des poissons avec une précision temporelle toute nouvelle. Un problème important dans l'analyse des données provenant des étiquettes récupérées est celui de la géolocalisation, c'est-à-dire de la détermination des déplacements du poisson en comparant les renseignements recueillis sur l'étiquette et les observations dans le milieu, par exemple de la température, de la marée, de la longueur du jour etc. Les résultats se présentent souvent sous forme de trajectoire, mais la variation de la précision dans l'espace et le temps est souvent considérable. Dans notre travail, nous utilisons un filtrage particulier pour déterminer le positionnement de morues franches (*Gadus morhua*) dans la Baltique, ce qui représente les résultats comme des distributions de probabilité à chaque échelon temporel, donnant ainsi une représentation explicite de l'incertitude. De plus, la méthode sert à estimer l'importance de l'erreur dans les mesures des étiquettes et dans la vitesse de nage des poissons. Nous estimons la vitesse moyenne de nage durant le jour à environ  $0.20 \text{ m}\cdot\text{s}^{-1}$  chez des poissons de  $\sim 60 \text{ cm}$  de longueur. La méthode est générale et elle est présentée de façon à être utilisée facilement dans des systèmes différents où d'autres variables à mesurer sont envisagées.

[Traduit par la Rédaction]

## Introduction

Observation of the behaviour of individual fish in situ is difficult and constrained to short periods of time. The understanding of the behaviour of fish is therefore much less developed than that for terrestrial animals, for example, for many fish stocks, not even the spawning location is known. The appearance of small electronic archival tags have therefore been a welcome new tool for obtaining information about fish. An archival tag is a small container mounted on the back of the fish or inserted into the body cavity that logs the state of the environment, e.g., pressure (depth), light, temperature, or salinity, on a regular basis. When the fish is caught, the data from the tag are recovered and provide in-

formation about the detailed behaviour of the fish over long time spans.

The data from the recovered tag contain information about vertical behaviour that can be used to infer changes in diurnal behaviour (Righton et al. 2001) or tune models of adjustments of the swimming bladder (Godø and Michalsen 2000). Temperature has been used in relation to the movement of fish across fronts (Stensholt 2001). Much effort is invested into using data from tags to geolocate the fish, i.e., to estimate the movement. The geographical variation in day length and the timing of sunrise or sunset can be used for fish travelling distances of oceanic scale, e.g., tuna (Sibert et al. 2003). For demersal fish like cod and flatfish living in seas with tidal movements, the amplitude and phase of the

Received 29 June 2006. Accepted 1 February 2007. Published on the NRC Research Press Web site at [cjfas.nrc.ca](http://cjfas.nrc.ca) on 28 April 2007.  
J19393

**K.H. Andersen,<sup>1</sup> U.H. Thygesen, and S. Neuenfeldt.** Danish Institute for Fisheries Research, Technical University of Denmark, Charlottenlund Slot, Jægersborg Allé 1, DK-2920 Charlottenlund, Denmark.

**A. Nielsen.** University of Hawaii at Manoa, 1000 Pope Road, MSB 313, Honolulu, HI 96822, USA.

**H.-H. Hinrichsen.** Leibniz-Institute of Marine Research, Düsternbrooker Weg 20, D 24105 Kiel, Germany.

<sup>1</sup>Corresponding author (e-mail: [kha@difres.dk](mailto:kha@difres.dk)).

tide can be inferred from the pressure changes. If there is a sufficient spatial variability in the tide, this can be used for geolocation (Hunter et al. 2003a, 2003b). Finally, variation in the fields of temperature or salinity can be used by comparing the recorded data from tags with circulation models (Nielsen et al. 2006; Neuenfeldt et al. 2007). Common for all the methods is the need to cope with large and spatial variations in uncertainties, governed by variations in the gradients in the observed fields (Hunter et al. 2003a). A proper geolocation, therefore, not only determines the most probable position of the tagged fish, but also the uncertainty associated with the geolocation. Explicit calculation of the uncertainty is a major focus of this study. Information from geolocated fish provide valuable information needed for spatial explicit management of fish stock, like the closure of specific areas for fishing (marine protected areas, MPAs). If many fish are geolocated, the information can be used to discriminate among different local stocks or analyze the use of MPAs.

The methods used for geolocation can roughly be partitioned into heuristic and state-space-based methods. In the heuristic methods, all the possible positions of a tag at a given time is determined from the environmental observations. From these, the position with the best fit is selected inside a range close to the current position of the tag, determined by the maximum velocity of the fish (e.g., Domeier et al. 2005; Neuenfeldt et al. 2007). State-space-based methods (most prominently the Kalman filter) are well known statistical methods used for tracking objects on radar (Harvey 1989). They have been used for wildlife telemetry (Anderson-Sprecher and Ledolter 1991), for geolocating simulated turtles using Markov chain Monte Carlo (MCMC) (Jonsen et al. 2003), and for geolocating tuna using the Kalman filter (Sibert et al. 2003). Roughly speaking, they estimate the most probable track that behaves according to a basic movement model, typically a random walk, and corresponds to the environmental observations from the tag. The simplest state-space-based method is the Kalman filter, which assumes that the errors, e.g., the differences between the observations from the tag and the modelled data (e.g., day length), are Gaussian distributed and yield a Gaussian probability density function (pdf) for the position. This may be a reasonable approximation for light- or tidal-based methods far from land, but when the tag is close to land or the fields that the tag is compared with are complex, this assumption is no longer reasonable. This assumption can be relaxed by using the particle filter (Ristic et al. 2004), which has been used for geolocation on synthetic data (Nielsen 2004; Royer et al. 2005). A strong feature of the state-space-based methods is that it is possible to estimate the parameters entering into the movement models (Jonsen et al. 2003; Sibert et al. 2003). In the case of a random walk model, this would be the effective diffusivity of the fish.

In this work, we will use the particle filter to geolocate Atlantic cod (*Gadus morhua*) in the Baltic Sea based on salinity, bathymetry, and recatch location. Moreover we will demonstrate how not only the most probable value of the parameters can be estimated, but their whole posterior distribution. We will also show how a proper assessment of uncertainty in the geolocation is made by combining the uncertainty on the geolocation method itself with the uncer-

tainty with which the parameters are estimated. The methodology is general and can be applied to geolocate fish using any available observations: light, tides, bathymetry, hydrography, recatch location, etc. Finally, we will show how the geolocations can be used to infer the usage of a MPA by the tagged fish.

## Materials and methods

The aim of this methodological description is to provide a readily useable recipe for using the particle filter for geolocation and estimate model parameters. Therefore, rigor will in some cases be sacrificed for clarity. For a more stringent introduction to the particle filter for geolocation, see Royer et al. (2005).

The result of the geolocation is a pdf describing the possible positions of the tagged fish for each time  $p^n(\mathbf{x})$ , where the superscript  $n$  refers to the discrete time step, and  $\mathbf{x}$  is the position vector (in the horizontal plane:  $\mathbf{x} = (x_{\text{longitude}}, x_{\text{latitude}})$ ). In the application of a state-based method, the calculation of  $p$  is done in two steps: (i) a forward in time estimation, where the pdf at time  $n$  is based on the pdf from the previous step and on measurements at the current step  $\mathbf{y}^n$ . Effectively the pdf is therefore calculated on the basis of all previous measurements:  $p^n(\mathbf{x}|\mathbf{y}^0 \dots \mathbf{y}^n)$  and (ii) a smoothing step where information from the whole time series is taken into account at each time step  $p^n(\mathbf{x}|\mathbf{y}^0 \dots \mathbf{y}^N)$ , where  $N$  is the number of steps. The pdf during the forward estimation is constructed by Monte Carlo simulation, such that it is the sum of a large number of delta distributions or particles. The particles are simulated based on a movement model and resampled using weights calculated from the measurements  $\mathbf{y}^n$ . This basically constitutes a standard particle filter (Doucet et al. 2001). The smoothing step is a backward calculation where the particles are also resampled backwards in the calculation. The pdf of the tag positions depends on a number of parameters  $\boldsymbol{\theta}$ . By defining a likelihood of the whole pdf as a function of the parameters  $\ell(\boldsymbol{\theta})$ , the value of the parameters can be estimated by maximizing the likelihood. Finally, the error estimates of the tag position from the pdf is corrected by taking the precision of the estimation of the parameters into account, thereby increasing the error estimate.

## Movement model

The motion of the individual particles is described by a movement model that corresponds to the basic movement of the individual fish. The usual assumption is that the basic movement is a random walk with Gaussian increments. This has the disadvantage that there is a finite probability for an arbitrarily large movement in finite time. As fish typically have an upper limit for their maximum velocity, a movement model is proposed where the new position is given by an uniform random distribution within a radius given by the maximum velocity  $U$  of the fish. Thus, the predicted position of the  $i$ th particle in the  $n$ th time step  $\tilde{\mathbf{x}}_i^n$  is given by

$$(1) \quad \tilde{\mathbf{x}}_i^n = \mathbf{x}_i^{n-1} + U \Delta t \sqrt{\xi_i} \begin{bmatrix} \cos(\psi_i) \\ \sin(\psi_i) \end{bmatrix} \quad \text{where } \psi_i = 2\pi\zeta_i$$

where  $\Delta t$  is the time step, and  $\xi_i$  and  $\zeta_i$  are drawn from a uniform distribution between 0 and 1. For times larger than

$\Delta t$ , this movement model corresponds to a random walk. For areas with strong currents, these could be added to the movement model.

### Error model

Particles are associated with a weight  $w_i^n$ , calculated from the likelihood of the observation vector from the tag  $y^n$ , given the predicted position  $\tilde{x}_i^n$  of the particle. For Baltic cod, the observation contains three pieces of information: salinity ( $y_{\text{sal}}^n$ ), depth ( $y_z^n$ ), and for some of the measurements also an estimation of the bottom depth ( $y_{\text{depth}}^n$ ).

The tag used in this study contain a calibration bias, which is well modelled by a linear dependence on the salinity, such that the real salinity  $s$  is given by the measured one  $y_{\text{sal}}$  as  $s = A_{\text{sal}} + B_{\text{sal}} y_{\text{sal}}$ . The error from the hydrodynamic model is harder to model, and it is therefore assumed to be normally distributed, with zero mean and a standard deviation  $\sigma_{\text{sal}}$  independent of time. Thus, the contribution to the weight from the salinity is

$$(2) \quad w_{i,\text{sal}}^n = \phi(\Delta s; 0, \sigma_{\text{sal}})$$

where  $\phi(x; \mu, \sigma)$  is a Gaussian density with mean  $\mu$  and standard deviation  $\sigma$  for the value  $x$ .  $\Delta s$  is the difference between the salinity measured by the tag, corrected for the calibration and the salinity in the model at the position of the particle and the depth recorded by the tag ( $\Delta s = s(\tilde{x}_i^n, y_z^n) - (A_{\text{sal}} + B_{\text{sal}} y_{\text{sal}}^n)$ ). In practice, it turned out to be more convenient to use the bias in salinity at two fixed salinities instead of the parameters  $A_{\text{sal}}$  and  $B_{\text{sal}}$ , such that  $A_{\text{sal}} = \Delta_5 - 5$  psu and  $B_{\text{sal}} = (\Delta_{20} - \Delta_5)/(15 \text{ psu})$ , where  $\Delta_5$  and  $\Delta_{20}$  are the biases at 5 and 20 psu, respectively.

To calculate the contribution to the importance weights from the location of the bottom, it is assumed that the error on the bathymetry  $z_b$  is normally distributed with standard deviation  $\sigma_b$ . In the cases where the bottom depth  $y_{\text{depth}}^n$  can be inferred from the tag, this gives a weight

$$(3) \quad w_{i,\text{depth}}^n = \phi(\Delta z; 0, \sigma_b)$$

where  $\Delta z = z_b(\tilde{x}_i^n) - y_{\text{depth}}^n$  is the difference between the bottom depth from the bathymetry and the distance below the surface of the fish registered by the tag. At the times where the bottom depth cannot be estimated, we still know that the bottom must be below the fish, i.e.,  $\Delta z$  cannot be (very) negative. The weight should therefore be small if the depth of the tag is below the bathymetry and large if it is above. Using the same error model for the bathymetry as above, the weight becomes

$$(4) \quad w_{i,\text{depth}}^n = \mathcal{N} \int_{-\infty}^{y_{\text{depth}}^n} \phi(\Delta z; 0, \sigma_b) dz$$

with the normalization  $\mathcal{N}$  being calculated such that the integral of  $w$  from  $-\infty$  to the surface is 1. This can be formulated with error functions as

$$(5) \quad w_{i,\text{depth}}^n = [1 + \text{erf}(\Delta z)] \times \left\{ z_b \left[ 1 + \text{erf} \left( \frac{z_b}{\sqrt{2} \sigma_b} \right) \right] - \sqrt{\frac{2 \sigma_b^2}{\pi}} e^{-z_b^2/(2\sigma_b^2)} \right\}^{-1}$$

The total importance weight is the product of the different contributions:

$$(6) \quad w_i^n = w_{i,\text{sal}}^n w_{i,\text{depth}}^n$$

which is normalized to form

$$(7) \quad \hat{w}_i^n = w_i^n / \sum_i w_i^n$$

The importance weights will be used to resample the particles, such that those with higher weights are preferred over those with lower weights.

### Particle filter

The procedure used for simulating the particles in the particle filter is the following:

- Step 0: The filter is initialized by setting all the  $N_p$  particles to the deployment point, which is assumed to be known without error.
- Step 1: Simulate the movement of each particle using eq. 1. If the particle is on land, a new position is tried.
- Step 2: Calculate the weights  $\hat{w}_i^n$  for each particle using eq. 7.
- Step 3: Resample the particles with replacement with a probability proportional to their weight.
- Step 3a: Form an index  $\mathcal{I}^n$ , where each entry refers to a particle number before resampling, with probability

$$(8) \quad P(\mathcal{I}_i^n = j) = \hat{w}_j^n$$

In practice this is done by first forming the cumulative sum of the weights  $F_j$ . Then, for each entry of  $\mathcal{I}^n$ , choose a random number  $\xi_i$  between 0 and 1 and find the particle  $j$  for which

$$(9) \quad \mathcal{I}_i^n = \min_j \{F_j \geq \xi_i\}$$

- Step 3b: Resample the particles at the current *and previous* time steps with replacement using the index  $\mathcal{I}^n$  such that the position of the  $i$ th particle at time  $n$  and resampling  $m$ :  $\mathbf{x}_i^{n,m}$  is

$$(10) \quad \mathbf{x}_i^{n,n} = \tilde{\mathbf{x}}_{\mathcal{I}_i^n}^n$$

$$(11) \quad \mathbf{x}_i^{n-l,n} = \mathbf{x}_{\mathcal{I}_i^n}^{n-l,n-1} \quad \text{where } l \in [1 \dots (n-1)]$$

This resampling of the particles effectively weeds out those with lower weights and replaces them with particles with higher weights.

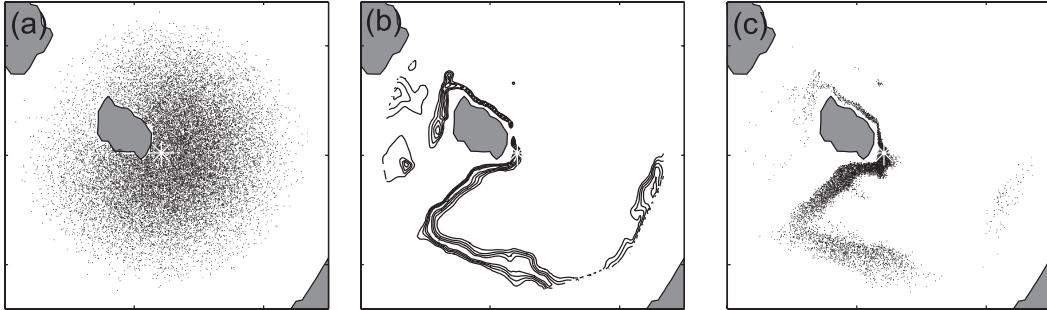
Steps 1 to 3 are iterated for all time steps (Fig. 1). At the final time step information about the recatch position  $\mathbf{x}_{\text{catch}}$  can be included in the calculation of the weights:

$$(12) \quad w_i^N = \phi(|\mathbf{x}_{\text{catch}} - \mathbf{x}_i^N|; 0, \sigma_{\text{catch}})$$

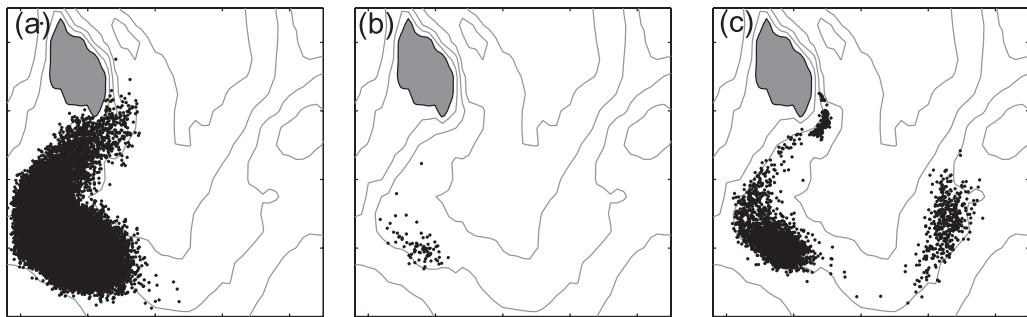
where  $\sigma_{\text{catch}}$  is an approximate precision with which the recatch position is known. This final step is important as it contains much information, which is propagated back into the track during resampling. It is also crucial information for the estimation of parameters.

The particles  $\mathbf{x}^{n,m}$  represents the pdf of the position of the tag at time  $n$  based on information from the times steps  $1 \dots m$ . This means that in the end we have  $\mathbf{x}^{1 \dots N,N}$ , which is the representation of the pdf of the location of the fish based

**Fig. 1.** Illustration of the steps in the particle filter: (a) The particles 3 days after release from the white star, without resampling. (b) Contour plot of the weights calculated from the salinity and the bathymetry. At this point the tag is close to a bottom contour line. (c) The particles left after resampling, based on the information at the current time step.



**Fig. 2.** Illustration of the distribution of particles at day 145 for tag 2 at different stages in the algorithm. (a) All  $2^{16}$  particles at day 145. (b) The particles at day 145 at the end of the calculation, when the particles have been resampled by the time steps from day 145 and onwards. (c) Particles from 50 realizations with random parameters. Note how the particles are clustered around the 45 m contour line, which is the estimated bottom depth at this day.



on information from the whole time series. In this manner, information is passed all the way back into the previous calculated particles at each time step (Figs. 2a, 2b). In the pure particle filter (Ristic et al. 2004), only the particles at the current time step are resampled. The resampling of all the time steps constitutes what in the milieu of the Kalman filter is known as a smoothing step, during which information is propagated all the way from the first time step to the last and back again (Harvey 1989). It has been shown that the particle filter parallel to the smoothing can be obtained exactly by the resampling of the particles backwards in time (Kitagawa and Sato 2001). The resampling is not without consequences, as the number of different particles declines during each resampling. This leads to a diluted representation of the pdf, in particular in the beginning of the track, where the particles are resampled most times. This can be compensated by having a large number of particles, but if the track contains many time steps, this might not be feasible. In that case, the resampling in eq. 11 can be reduced to a finite number of steps back in time, which usually resembles the global estimate (Kitagawa and Sato 2001).

### Estimation of parameters

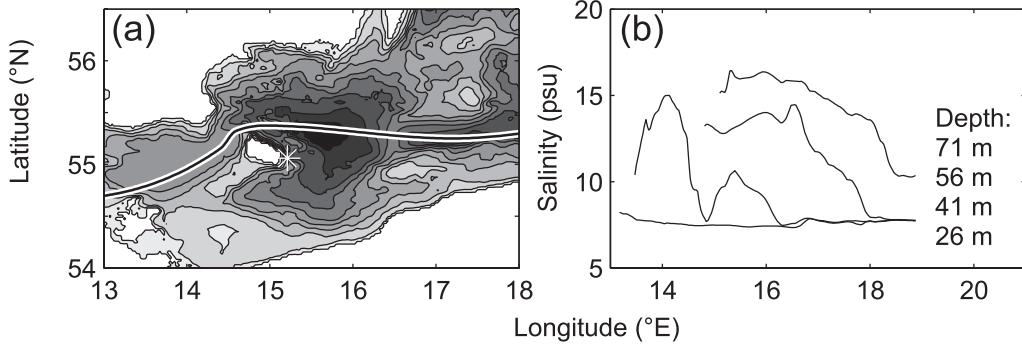
The parameters of the model are the maximum velocity, the standard deviation on the salinity, the two parameters for the correction of the salinity, and the standard deviation on the bathymetry:  $\Theta = (U, \sigma_{\text{sal}}, \Delta_5, \Delta_{20}, \sigma_b)$ . With the Kalman filter, these parameters can be estimated from the tag data by a maximum likelihood approach (Sibert et al. 2003). We use another likelihood-based method, namely Bayesian maxi-

mum posterior probability, which has the advantage that we not only calculate the most probable value of the parameters but also the distribution of their values, and it is possible to take prior information into account. Additionally, the temporal correlation of the measurements are taken into account. Assuming that the measurements  $y$  are uncorrelated in time, an approximation of the negative log-likelihood for an estimated track given the parameters  $\Theta$  is (Higuchi 2001; Kitagawa and Sato 2001)

$$(13) \quad \ell(\Theta) = -\log L(y|\Theta) = -\sum_{n=1}^N \log \overline{w^n}$$

where  $\overline{w^n}$  is the average of the un-normalized weights of each particle  $w_i^n$ . The error on the measurements in eqs. 2 and 3 are assumed to be uncorrelated between measurements, but in reality they are strongly correlated, e.g., a measurement of salinity at one time step is quite similar to the one at the next time step, as the movement of the fish and the changes in salinity are typically slower than each measurement. A measure of the correlation can be obtained by assuming that the fish does not move, such that the auto-correlation of, e.g., salinity in the area  $c(t)$  can be used. The correlation time  $\tau$  can then be calculated as  $\tau = \int_{-\infty}^{\infty} c(t) dt$ . As each observation is not independent of the previous one, this should be reflected in the calculation of the likelihood. The actual number of independent observations is equal to the number of observations  $N$  divided by the estimated correlation time  $\tau$ . This is taken into account by dividing the nega-

**Fig. 3.** (a) Contours of the bathymetry in the study area with 20 m equidistance. The island in the middle is Bornholm. The white star just east of Bornholm is the tagging station. (b) Salinity at four different depths along the white line in the panel *a*, at the day when the fish were tagged.



tive log-likelihood with the correlation time  $\tau$ , measured in number of time steps:

$$(14) \quad \ell(\boldsymbol{\theta}) = -\sum_{n=1}^N \log(\overline{w^n})/\tau$$

This solution to the problem is not a fully correct way of including correlated errors. However, in this situation we do not know what the exact correlation structure is, nor can we estimate it. What we do know is that correlated errors will degrade the performance of the estimators. The solution we have employed is motivated by the problem of estimating a parameter  $\mu$  based on noisy observations  $X_i = \mu + E_i$ , where  $E_i$  are correlated, with a correlation time that is small compared with the number of measurements. For this situation, the estimator itself does not change with the autocorrelation, but its variance is proportional to the correlation time, so that the likelihood is inversely proportional to the correlation time, just as we have done in eq. 14. Regarding the form of the autocorrelation function of the error, the most important contribution is model error in the circulation model. Within the field of circulation modelling, little is known about the correlation structure of model errors. In this situation, equating the correlation time of the error with that of the signal is to be seen as a first estimate. The best argument supporting it is really that we have no other characteristic time scales at hand, and neglecting the autocorrelation will result in a grossly over-optimistic picture of the uncertainty of the estimates.

Estimating the parameters  $\boldsymbol{\theta}$  by optimizing the likelihood with a standard Newton-type optimizer turned out to be difficult and unstable. Furthermore, the standard approximation of the covariance matrix based on the curvature of the likelihood turned out to be useless. This is because the likelihood function itself is estimated from a number of simulated particles. The usual trick for optimizing likelihood functions based on simulations is to fix the random seed, or in other words, reuse the random numbers that the likelihood relies on every time the likelihood is calculated. This is not sufficient in our case, as the resampling in step 3 depends on the parameter values. These two steps introduce small discontinuities in the likelihood function, because even an infinitesimal small change in the parameter values can cause a finite jump in the likelihood. These discontinuities are reduced if the number of particles are increased, but it would require a

huge amount of particles (and computation time) to reduce the size of these discontinuities below the level of the numerical precision required to reliably approximate the curvature of the likelihood by a finite difference approximation. Instead the following MCMC approach is suggested (Nielsen 2004). The algorithm is the Random Walk Metropolis–Hastings (RWMH) (Gilks et al. 1996). In this algorithm, an estimated parameter vector  $\boldsymbol{\theta}$  makes a random walk in parameter space. For each new step, the likelihood of the proposed step is tested, and if it is sufficiently better than the current estimate, this step is accepted. In this way, the random walker is being confined to the part of parameter space with highest likelihood.

RWMH step 0: Initialize the parameters.

$$(15) \quad \boldsymbol{\theta}^0 = (U, \sigma_{\text{sal}}, \Delta_5, \Delta_{20}, \sigma_z)$$

$$(16) \quad = (0.25 \text{ m}\cdot\text{s}^{-1}, 0.5 \text{ psu}, 0 \text{ psu}, 0 \text{ psu}, 5 \text{ m})$$

RWMH step 1: Sample a proposal parameter vector at iteration step  $l$ :  $\boldsymbol{\theta}'^l$  from a normal distribution, with mean  $\boldsymbol{\theta}^{l-1}$  and fixed variances, specified as  $\sigma = 0.04 \text{ m}\cdot\text{s}^{-1}, 0.2 \text{ psu}, 0.4 \text{ psu}, 0.4 \text{ psu}, 0.2 \text{ m}$ .

RWMH step 2: Calculate the likelihood  $\ell(\boldsymbol{\theta}'^l)$  using the particle filter and eq. 14. Note that in step 3b it is only necessary to resample the particles at the current time step.

RWMH step 3: Preferably accept the new proposal if the calculated likelihood is smaller than the previous ones. Specifically use

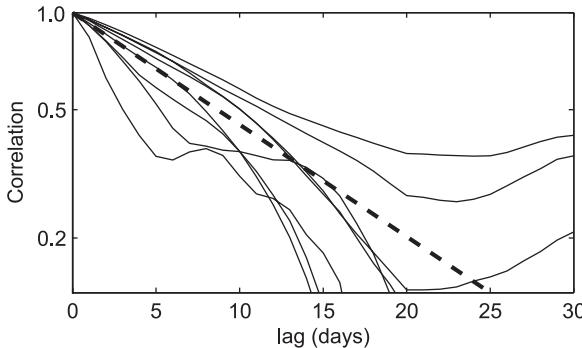
$$(17) \quad \boldsymbol{\theta}'^l = \begin{cases} \boldsymbol{\theta}'^l & \text{if } \mathcal{U}(0, 1) \leq \exp[\ell(\boldsymbol{\theta}^{l-1}) - \ell(\boldsymbol{\theta}'^l)] \frac{P(\boldsymbol{\theta}'^l)}{P(\boldsymbol{\theta}^{l-1})} \\ \boldsymbol{\theta}^{l-1} & \text{otherwise} \end{cases}$$

where  $P(\boldsymbol{\theta}'^l)$  is a prior distribution incorporating other known information about the possible value of the parameters.

RWMH step 4: Increase the iterator  $l$  and go to step 1, until  $L$  iterations.

An important practical note is that the negative log-likelihood values in step 3 can not be reused, which implies that  $\ell(\boldsymbol{\theta})$  must be recalculated every time it is needed. If the values are reused, then the resulting chain tends to get stuck for very long periods at the same values. This is due to the

**Fig. 4.** Autocorrelation of the salinity in the centre of the Bornholm Basin ( $15.5^{\circ}\text{E}$ ,  $55.3^{\circ}\text{N}$ ) has been calculated at depths from 20 to 90 m. The broken line corresponds to an exponential correlation with a time constant of 12.5 days.



fact that  $\ell$  is simulated. Once in a while  $\ell(\boldsymbol{\theta})$  randomly hits a very low value, and if not re-evaluated, this low value will reject new proposals for a long time. When the calculation is finished, a pdf of the value of the parameters can be constructed from the series  $\boldsymbol{\theta}$ . Prior information was used in the case of the calibration error on the salinity, which was restricted to be in the interval  $[-5:5]$  psu. The values of the other parameters were restricted to be only positive.

### Error on track

An estimate of the uncertainty of the geolocation at a given time can be found from the distribution of particles for a given set of the parameters  $\boldsymbol{\theta}$ :  $p^n(\mathbf{x}, \boldsymbol{\theta})$ . If the parameters are estimated with great accuracy, it would be sufficient to consider the pdf for the most probable values of the parameters. Since there is uncertainty in the estimation of the parameters, that has to be taken into account as well. This is accomplished by making a large number of simulations for random values of the parameters, taken from the RWMH series  $\boldsymbol{\theta}^l$ . For each of the randomly chosen parameter values  $\boldsymbol{\theta}$ , the pdf  $p^n(\mathbf{x}, \boldsymbol{\theta})$  for the position of the tag is calculated. The final pdf is constructed by averaging all the pdfs:  $p^n = p^n(\mathbf{x}, \boldsymbol{\theta})$ . This corresponds to the standard way of generating confidence intervals for predictions based on estimated regression models and also has a Bayesian interpretation in terms of the posterior distribution of the position given the measurements. The effect of including the error on the estimated parameters in the final pdf is that the particles from many different runs with different parameters values cover a much larger area than the run with just one set of parameter values (Fig. 2).

## Results

To demonstrate the applicability of the particle filter for geolocation of tagged fish, it will be applied to two data sets obtained from tagged cod in the Baltic Sea. Information of the salinities are obtained from a hydrodynamic model on a  $5\text{ km} \times 5\text{ km}$  grid every day and at 3 m depth intervals (Neuenfeldt et al. 2007).

**Table 1.** Summary of estimated parameters from the tags, shown as the mean  $\pm$  1 standard deviation.

Tag No.	Days out	Velocity ( $\text{m}\cdot\text{s}^{-1}$ )	$\sigma_z$ (m)	$\sigma_{\text{sal}}$ (psu)
1	144	$0.035 \pm 0.051$	$0.87 \pm 0.37$	$0.84 \pm 0.78$
2	105	$0.26 \pm 0.30$	$1.3 \pm 0.77$	$3.9 \pm 1.2$
3	126	$0.054 \pm 0.048$	$1.2 \pm 0.45$	$0.60 \pm 0.91$
4	108	$0.13 \pm 0.12$	$5.3 \pm 3.8$	$2.4 \pm 1.1$
5	144	$0.34 \pm 0.36$	$5.1 \pm 4.6$	$4.0 \pm 2.5$
6	77	$0.29 \pm 0.13$	$5.9 \pm 4.0$	$2.6 \pm 3.2$
7	154	$0.20 \pm 0.087$	$5.7 \pm 4.6$	$3.3 \pm 1.2$
8	67	$0.21 \pm 0.14$	$13 \pm 7.8$	$10 \pm 7.7$
9	54	$0.24 \pm 0.14$	$9.4 \pm 5.5$	$3.0 \pm 2.0$
10	158	$0.21 \pm 0.14$	$7.8 \pm 4.0$	$1.9 \pm 0.84$
Mean	—	$0.20 \pm 0.15$	$3.3 \pm 2.1$	$5.6 \pm 3.6$

### Tagging experiments

The cod were tagged in April 2003, just east of the island Bornholm (Fig. 3a). The average length of the tagged fish was 56 cm, and the two examples shown here were 59.5 and 57.6 cm. East of the tagging station is the Bornholm Basin, a circular basin with a maximum depth of  $\sim 100$  m. Further east, at  $\sim 16.5^{\circ}\text{E}$  longitude, is the Stolpe Trench, forming a sill into the Gdańsk Basin, located to the east of the map. The driving part of the geolocation is the gradients in salinity, which are most pronounced in the deeper waters (Fig. 3b).

The tags measure pressure, salinity, and temperature, with a temporal resolution between 5 and 15 min. To simplify the example, information from the temperature has been disregarded. As temperature is highly correlated to salinity in the deep water, the temperature does not contain much additional information.

To estimate the correlation time  $\tau$ , the autocorrelation of salinity in the centre of the Bornholm Basin has been calculated at different depths (Fig. 4). It is seen that correlation is roughly exponential,  $c = e^{-t/T}$ , with a time constant  $T$  of  $\sim 12.5$  days. The correlation time can be estimated as  $\tau = \int_{-\infty}^{\infty} c dt = 2T$  or  $\sim 25$  days.

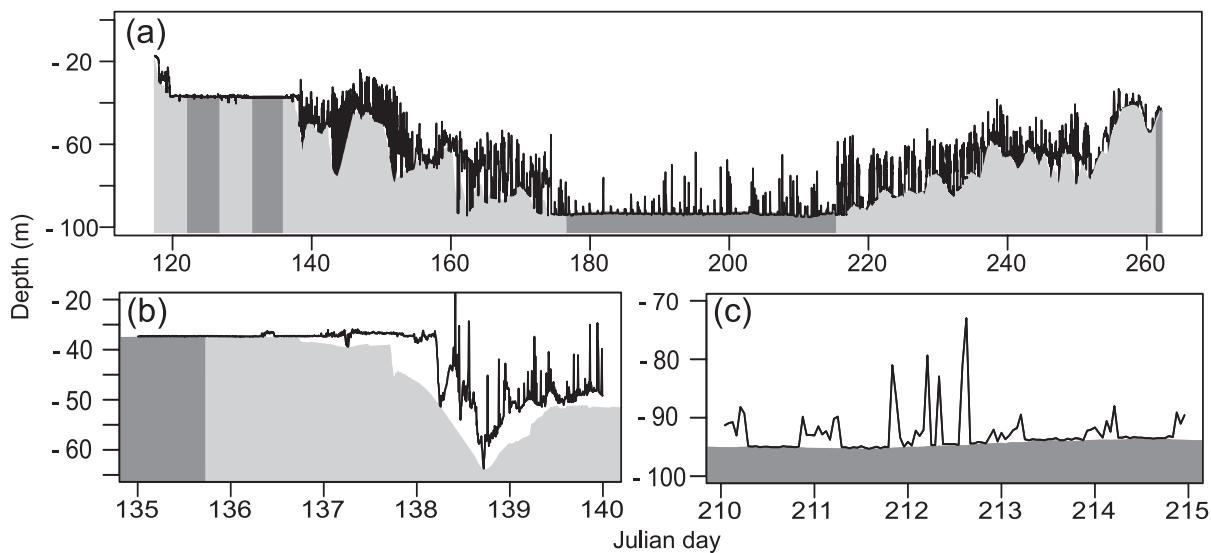
The time step used in the calculation was 1 day. The calculation of the most likely set of parameters was done using  $2^{12}$  particles and 15 000 iterations, discarding the first 1000 iterations. For the final calculation of the pdfs,  $2^{16}$  particles were used for each of 50 runs with random parameter values. The algorithm was programmed in MATLAB (The MathWorks, Inc., Natick, Massachusetts) and run on a 3 GHz standard PC. A run with 200 days takes  $\sim 12$  h for estimating the parameters and 2 h for the 50 realizations.

Ten tags that have been out from 54–158 days have been geolocated and the parameters estimated (Table 1). The maximum movement velocity varied between 0.04 and  $0.33\text{ m}\cdot\text{s}^{-1}$ , with a mean of  $0.20 \pm 0.15\text{ m}\cdot\text{s}^{-1}$ .

### Track 1

Just after the release, the fish seeks a depth of a little less than 40 m and seems to stay in the same position without much vertical activity for a fortnight (Fig. 5). At day 138, it moves deeper in the water column. This continues until it

**Fig. 5.** The depth recorded by the tag over the period from the release of the fish till it was caught. The light-shaded area indicates the estimation of the minimum bottom depth. The areas with the darker shading is when it was assumed that the minimum depth was equal to the bottom depth. Also shown are the initiation of migration (*b*) and examples of activity at night and inactivity during daytime (*c*).



reaches a depth of 90 m. Here it stays for longer than a month, after which it moves to even lower depths, until it is eventually caught. A typical observation from the time series of the depth is that the fish frequently go to a fixed depth level and then make excursions upwards in the water column. A likely interpretation is that the fish dwells at the bottom and then feeds up in the water column. This fits well with cod being a predominantly demersal species, often feeding on pelagic species, in this case herring and sprat. This is used to estimate the position of the bottom in the periods when the fish displays this behaviour. For each measurement point, a guess of the possible location of the bottom was made by finding the deepest point among the measurements in the range  $\pm 12$  h. If a large part of the measurements is close to this depth, this depth can be assumed to be the bottom. In practice, the histogram of the measurements in the range was generated, and if more than 15% of the measurements were within 10% of the depth range to the deepest point, it was assumed that this point was the location of the bottom. This criterion was chosen conservatively, such that it did not make any erroneous identifications of the bottom (Fig. 5).

The pdfs of the parameters show an interesting bimodal distribution, most prominently in the errors on the depth  $\sigma_z$  and salinity  $\sigma_s$ , with small secondary peaks at high values of the errors (Fig. 6). These show up in the plot of the final distributions of possible locations as a finite probability that the fish is going to the eastern basin (Fig. 6). An error in the hydrographic model on the salinity of 10 psu and on the bathymetry of 10 m is unlikely, and we can therefore exclude this possibility. This could also have been achieved by constraining the values of these parameters by further priors in eq. 17.

## Track 2

The calculations for this track were performed as for track 1 (Fig. 7). This example is more straightforward and predicts

that the fish stay close to the release position, probably with a short southwards migration.

## Population consequences

Having an explicit representation of the movement of a representative number of fish from a population, it is possible to calculate the habitat use of the population as a whole. However, having tagged fish in only one position, as in the current case, does not provide a representative sample of the population. Rather, it is a sample of the part of the population, which at the time of tagging was residing in the area where the tagging took place. Keeping that in mind, a relevant question for Atlantic cod in the Baltic Sea is what effect an enforced MPA had on the population. We have calculated the pdf of all the  $M = 10$  fish as

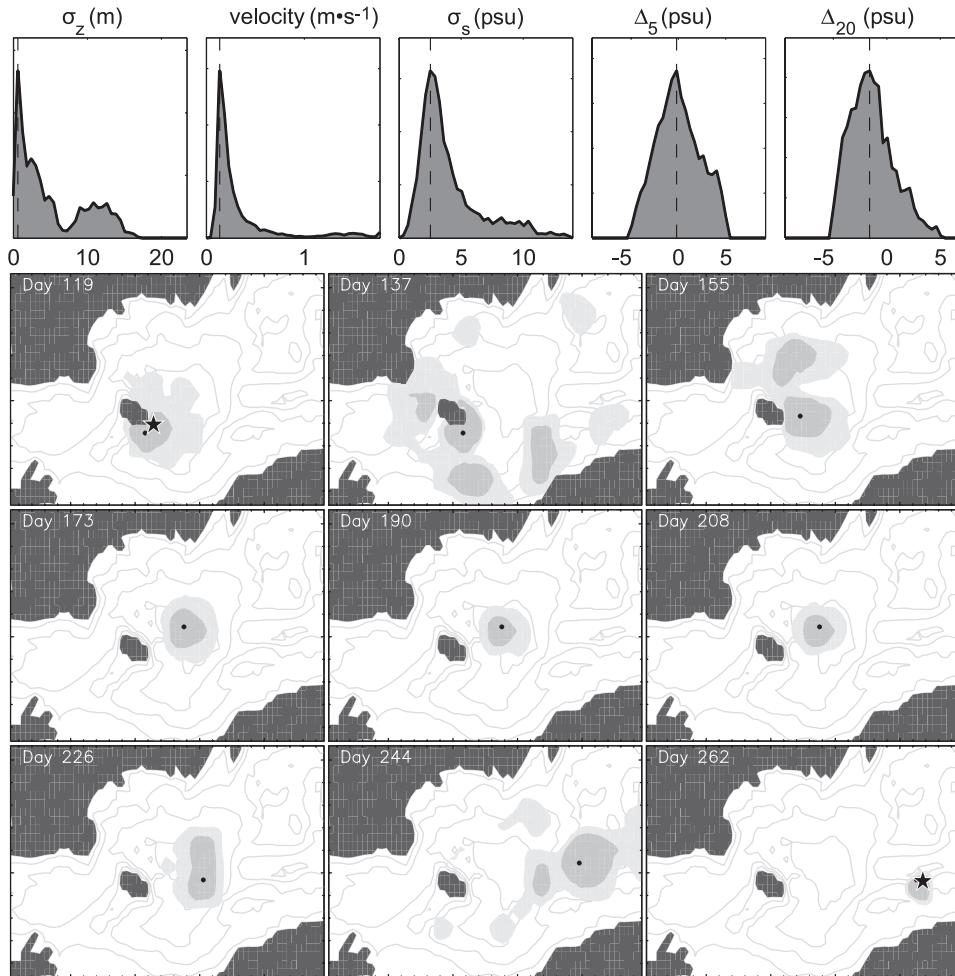
$$(18) \quad p_{\text{MPA}}(\mathbf{x}) = \frac{1}{M} \sum_{j=1}^M \int p_j(\mathbf{x}, t) dt$$

where the time integral is over the time period when the MPA was enforced. This shows that on average, the MPA was protecting around 10% of the tagged population (Fig. 8). To be more efficient, the MPA should be increased in the southerly direction, and the time period should be shifted to later in the season.

## Discussion

With the introduction of the particle filter into the toolbox for analysis of data from electronic tags, the ability to make accurate geolocation has been greatly improved. We have made a practical introduction on how to apply the particle filter, with particular focus on making a correct estimation of the error on the geolocation. This is particularly important for cases where the error structure is not simple, e.g., when the tags are geolocated using hydrographic data, when the tags are close to land, or when there are substantial uncertainty in the parameters entering into the geolocation. Apart

**Fig. 6.** The estimated parameters and probability density function (pdf) of track 1. The pdf is formed by assembling the particles from 50 realizations of the particle filter with random parameter values taken from the calculated distribution, including cross correlations. The light and dark grey contours are 95% and 66% confidence intervals, respectively. The stars are the catch–release positions, and the black dot is at the point with highest probability.



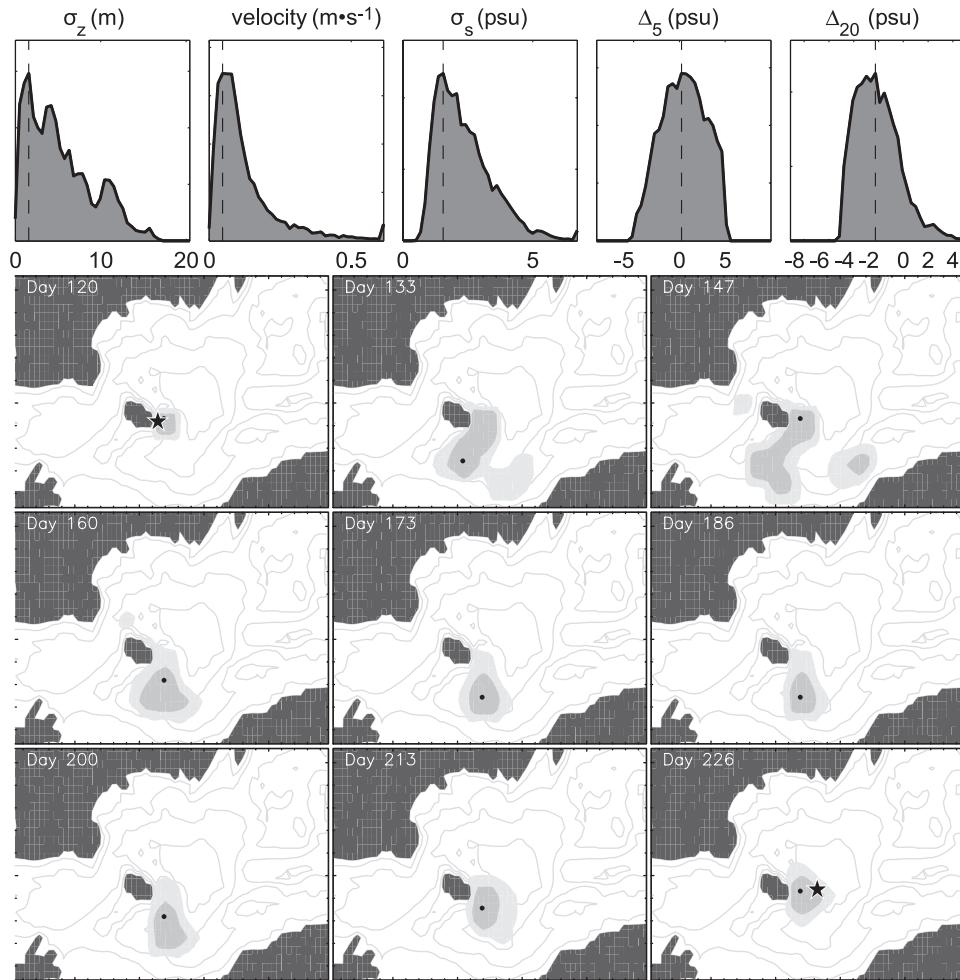
from the evaluation of the error, we have shown how to take information from the bathymetry into account. This is valuable information, in particular when the bottom depth can be confidently estimated. This will be of particular importance for geolocation of demersal species like flounder, rays, and skates. The method for determining the bottom depth presented here is fairly heuristic, and therefore the parameters were chosen conservatively, to avoid false estimations. In tidal regions, the sinusoidal tidal signal will be evident in the pressure signal when the fish is at the bottom (Metcalfe and Arnold 1997). In these cases, the bottom depth can be confidently determined, rendering the method more powerful. In areas with strong tidal currents, fish have been shown to utilize these for transport by changing their vertical position in the water column in phase with the tidal cycle (Harden Jones et al. 1979). In that case, information about the current should be taken into account in the movement model.

The use of the particle filter for geolocation has been demonstrated on synthetic data previously (Royer et al. 2005). In this case, only a forward estimation was demonstrated, which mean that the method is not global in time.

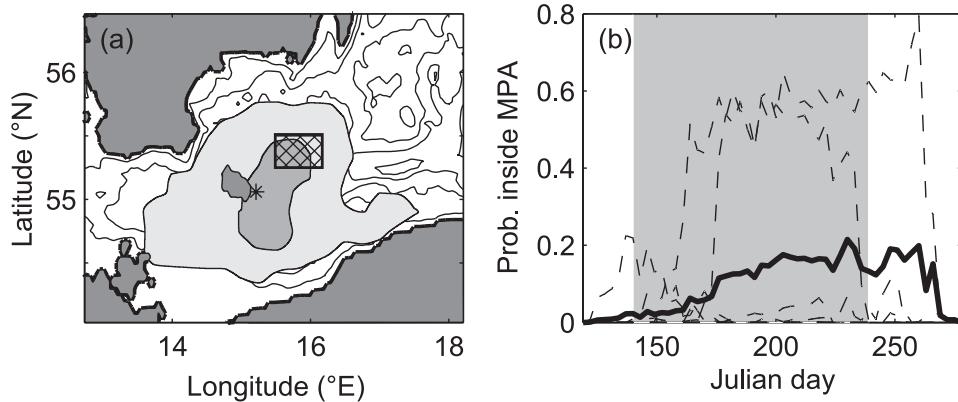
By also applying a smoothing step (step 3b), we have made a global estimation of the track. This weeds out dead ends in the estimated track, but more importantly it ensures that information from the catch location is propagated backwards through the time series, thus substantially improving the accuracy of the geolocation.

The method was demonstrated on tag data from Atlantic cod in the Baltic Sea, used previously for geolocation with a simpler method (Neuenfeldt et al. 2007). The particle filter specifically identified periods where the geolocation was very uncertain and periods where the position of the fish can be well established. A particular feature of this example is the use of the estimate of the bottom depth at times when it is available. In the Central Baltic Sea, there are frequent periods of anoxic conditions near the bottom in the deep basin, such that cod are unable to reach the bottom, and consequently no information about the bottom depth can be obtained. The two examples were from a year with favourable oxygen conditions, such that cod could reach the bottom at all times; however, still in that case, reliable estimates of the bottom depth were only found in less than half the time period. The calculated confidence intervals are rather large,

**Fig. 7.** The estimated parameters and probability density function (pdf) of track 2. The pdf is formed by assembling the particles from 50 realizations of the particle filter with random parameter values taken from the calculated distribution, including cross correlations. The light and dark grey contours are 95% and 66% confidence intervals, respectively.. The stars are the catch–release positions, and the black dot is at the point with highest probability.



**Fig. 8.** (a) The probability density function (pdf) of all the tags in the period from Julian day 135 to 240 as calculated from eq. 18. The hatched area is the marine protected area (MPA), where fishing was prohibited during that period. (b) The probability that a given fish was inside the MPA as a function of time for each fish (broken lines) and average (thick line). The shaded area is the time period when the MPA was enforced.



showing that it is not possible to geolocate fish accurately. However, it is clearly possible to identify in which basin the fish resides and also indicate in which part of the basin it is.

The estimated maximum daily average velocities were around  $0.20 \pm 0.15 \text{ m}\cdot\text{s}^{-1}$ . These velocities are similar to the theoretically estimated migration velocity, which is around  $0.3 \text{ m}\cdot\text{s}^{-1}$  for this size of fish (Ware 1978). However, some fish had smaller estimated movement rates. This indicates that either the fish do not migrate in a straight line or they do not migrate uninterrupted during a whole day. This also points to a possible weakness in the random walk movement model. In this model, it is most probable that the fish does not move at all. Within the scale of accuracy of the geolocation, this is probably true in periods where the cod are foraging in a limited area. As foraging may very well be the preferred mode of activity, the estimation of the movement velocity is dominated by this. A more accurate movement model would split the possible movement into two different modes: foraging and migration. The foraging mode is described by a random walk, whereas migration is a movement with a preferred direction and a higher mean velocity. This could be taken into account in the particle filter in a similar way as when aircrafts with different manoeuvring modes are tracked (Ristic et al. 2004). This was also done by Morales et al. (2004), however, without considering measurement error. Making the movement more advanced also means introducing more parameters. With the current data set, there is simply insufficient information available to reliably estimate more than one parameter in the movement model. It may be possible to apply such a model where the basic information is more accurate, i.e., fish geolocated using tides, eventually in combination with temperature (Nielsen et al. 2006).

## Acknowledgements

This study was partly funded by DG Fisheries of the EU under the CODYSEY Project 25 (contract code Q5RS-2002-00813) and partly by REX-I (the Danish Ministry of Food, Agriculture and Fishery).

## References

- Anderson-Sprecher, R., and Ledolter, J. 1991. State-space analysis of wildlife telemetry data. *J. Am. Stat. Assoc.* **86**(415): 596–602.
- Domeier, M.L., Kiefer, D., Nasby-Lucas, N., Wagschal, A., and O'Brien, F. 2005. Tracking Pacific bluefin tuna (*Thunnus thynnus orientalis*) in the northeastern Pacific with an automated algorithm that estimates latitude by matching sea-surface-temperature data from satellites with temperature data from tags on fish. *Fish. Bull.* **103**: 292–306.
- Doucet, D., de Freitas, N., and Gordon, N. (Editors). 2001. Sequential Monte Carlo methods in practice. Springer, New York.
- Gilks, W.R., Richardson, S., and Spiegelhalter, D.J. 1996. Markov Chain Monte Carlo in practice. Interdisciplinary statistics. Chapman & Hall, New York.
- Godø, O.R., and Michalsen, K. 2000. Migratory behaviour of northeast Arctic cod, studied by use of data storage tags. *Fish. Res.* **48**: 127–140.
- Harden Jones, F.R., Arnold, G.P., Greer Walker, M., and Scholes, P. 1979. Selective tidal stream transport and the migration of plaice (*Pleuronectes platessa* L.) in the southern North Sea. *J. Cons. Int. Explor. Mer.* **38**: 331–337.
- Harvey, A.C. 1989. Forecasting, structural time series models and the Kalman filter. Cambridge University Press, London.
- Higuchi, T. 2001. Self-organizing time series model. In *Sequential Monte Carlo methods in practice*. Edited by A. Doucet, N. de Freitas, and N. Gordon. Springer, New York. pp. 429–444.
- Hunter, E., Aldridge, J.N., Metcalfe, J.D., and Arnold, G.P. 2003a. Geolocation of free-ranging fish on the European continental shelf as determined from environmental variables. *Mar. Biol.* **142**: 601–609.
- Hunter, E., Metcalfe, J.D., and Reynolds, J.D. 2003b. Migration route and spawning area fidelity by North Sea plaice. *Proc. R. Soc. Lond. B Biol. Sci.* **270**: 2097–2103.
- Jonsen, I.D., Myers, R.A., and Flemming, J.M. 2003. Meta-analysis of animal movement using state-space models. *Ecology*, **84**(11): 3044–3063.
- Kitagawa, G., and Sato, S. 2001. Monte Carlo smoothing and self-organising state-space model. In *Sequential Monte Carlo methods in practice*. Edited by A. Doucet, N. de Freitas, and N. Gordon. Springer, New York. pp. 177–195.
- Metcalfe, J.D., and Arnold, G.P. 1997. Tracking fish with electronic tags. *Nature (London)*, **387**: 665–666.
- 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*, **85**(9): 2436–2445.
- Neuenfeldt, S., Hinrichsen, H.-H., Nielsen, A., and Andersen, K.H. 2007. Reconstructing migrations of individual cod (*Gadus morhua* L.) in the Baltic Sea by using electronic data storage tags. *Fish. Oceanogr.* In press.
- Nielsen, A. 2004. Estimating fish movement. Ph.D. thesis, Department of Mathematics and Physics, Royal Veterinary and Agricultural University, Copenhagen, Denmark.
- Nielsen, A., Bigelow, K.A., Musyl, M.K., and Sibert, J.R. 2006. Improving light-based geolocation by including sea surface temperature. *Fish. Oceanogr.* **15**(4): 314–325. doi:10.1111/j.1365-2419.2005.00401.x.
- Righton, D., Metcalfe, J., and Connolly, P. 2001. Different behaviour of North and Irish Sea cod. *Nature (London)*, **411**: 156.
- Ristic, B., Arulampalam, S., and Gordon, N. 2004. Beyond the Kalman filter. Particle filters for tracking applications. Artech House, Norwood, Mass.
- Royer, F., Fromentin, J.-M., and Gaspar, P. 2005. A state-space model to derive bluefin tuna movement and habitat from archival tags. *Oikos*, **109**: 473–484.
- Sibert, J.R., Musyl, M.K., and Brill, R.W. 2003. Horizontal movements of bigeye tuna (*Thunnus obesus*) near Hawaii determined by Kalman filter analysis of archival tagging data. *Fish. Oceanogr.* **12**(3): 141–151.
- Stensholt, B.K. 2001. Cod migration patterns in relation to temperature: analysis of storage tag data. *ICES J. Mar. Sci.* **58**: 770–793.
- Ware, D.M. 1978. Bioenergetics of pelagic fish: theoretical change in swimming speed and ration with body size. *J. Fish. Res. Board Can.* **35**: 220–228.

**Copyright of Canadian Journal of Fisheries & Aquatic Sciences is the property of NRC Research Press and its content may not be copied or emailed to multiple sites or posted to a listserv without the copyright holder's express written permission. However, users may print, download, or email articles for individual use.**