Ecological Modelling 221 (2010) 2008–2017

Contents lists available at ScienceDirect

Ecological Modelling

journal homepage: www.elsevier.com/locate/ecolmodel

A pioneer validation of a state-space model of vessel trajectories (VMS) with observers' data

E. Walker[a,b,*], N. Bez[a]

[a] IRD (Institut de Recherche pour le Développement), UMR 212 EME, Avenue Jean Monnet, BP 171, 34203 Sète, France
[b] CLS, 8-10 rue Hermès, Parc Technologique du Canal, 31520 Ramonville Saint-Agne, France

ARTICLE INFO

Article history:
Received 16 November 2009
Received in revised form 8 May 2010
Accepted 18 May 2010
Available online 18 June 2010

Keywords:
State-space model
Trajectory
Vessel Monitoring Systems (VMS)
Markov
Bayes
Validation
Tropical tuna purse-seiners
Observers' data

ABSTRACT

In the context of the expansion of animal tracking and bio-logging, state-space models have been developed with the objective to characterise animals' trajectories and to understand the factors controlling their behaviour. In the fisheries community, the electronic tagging of vessels commonly designated by Vessel Monitoring Systems (VMS) is developing and provides a new insight for the understanding, the analysis and the modelling of the trajectories of vessels and their prospecting behaviour. VMS data are thus a clue for the proper definition of fishing effort which remains a fundamental parameter of tuna stock assessments. In this context, we used the VMS (recording of hourly positions) of the French tropical tuna purse-seiners operating in the Indian Ocean to characterise three types of movement (states) on the VMS trajectories (stillness, tracking, and cruising). Based on empirical evidences, and on the regular frequency of VMS acquisition, this was achieved by the development of a Bayesian Hidden Markov model for the speeds and turning angles derived from the hourly steps of the trajectories. In a second phase, states were related to activities disentangling stillness into fishing or stop at sea. Finally the quality of the model performances was rigorously quantified thanks to observers' data. Confronting model prediction and true activities allowed estimating that 10% of the hourly steps were misclassified. The assumptions and model' choices are discussed, highlighting the fact that VMS data and observers' data having different time resolutions, the effective use of validating data was troublesome. However, without validation, these analyses remain speculative. The validation part of this work represents an important step for the operational use of state-space models in ecology in the broad sense (predators' tracking data, e.g. birds or mammals trajectories).

1. Introduction

State-space models are commonly used in ecology and associated fields to analyse tracking data (Royer et al., 2005; Gutenkunst et al., 2007), in mark-recapture framework (Ovaskainen, 2004; Gimenez et al., 2007) or in the stock assessment context.

Apart from their primary role in control and surveillance, Vessel Monitoring System (VMS) data have recently started to be used in the perspective of management plans like for instance for UK trawlers (Witt and Godley, 2007), or to estimate trawling effort allocation (Mills et al., 2007) or re-allocation after opening a marine protected area (Rijnsdorp et al., 2001). VMS data have also been simulated by Deng et al. (2005) to determine the VMS acquisition frequency required to describe trawl tracks in the prawn fishery of New Zealand. The ecological potential of such source of information is increasingly acknowledged especially in the case of open ocean (and for highly migratory species), where no exhaustive scientific stock assessment survey can be undertaken. Moreover, some recent demonstrations highlighted that for fishing boats, and more specifically, for Peruvian anchovy purse-seiners, trajectories are those of efficient predators when the prey they are fishing are heterogeneous in space (Bertrand et al., 2005). Such considerations opened the use of the characteristics of the trajectories of vessels as a proxy for the abundance of the species they are targeting as this has been done for other marine top predators (Grémillet et al., 2004; Sims et al., 2006; Witt and Godley, 2007; Robinson et al., 2007).

However, to do so, it is necessary to relate the fishing activities and more precisely, the local behaviour of the vessel, to various levels of the prey's presence. Previous works on animals or fishers that ought to characterise trajectories as a whole (e.g. Levy flights by Mandelbrot, 1977; Viswanathan et al., 1996; Bertrand et al., 2005; Bertrand et al., 2007) or those that focussed on the sole identification of fishing operations (e.g. neural networks, Bertrand et al., 2008) or areas of restricted search (ARS) (Knoppien and Reddingius, 1985; Fauchald and Tveraa, 2003; Tremblay et al., 2007), did not

* Corresponding author at: IRD (Institut de Recherche pour le Développement), Avenue Jean Monnet, BP 171, 34203 Sète, France. Tel.: +33 4 99 57 32 19; fax: +33 4 99 57 32 95.
E-mail addresses: Emily.walker@ird.fr, Emily.walker@avignon.inra.fr (E. Walker).

0304-3800/$ – see front matter © 2010 Elsevier B.V. All rights reserved.
doi:10.1016/j.ecolmodel.2010.05.007

aim to qualify all the segments of a trajectory, segment by segment. The idea here was to build a model that associates one particular activity to any single element of a VMS trajectory. We identified four main activities: fishing, stopping, tracking and cruising. Other activities are described in the tuna fisheries literature (Gaertner et al., 1999). Indeed, the main objective when estimating the activity of the vessel for each segment of the trajectory was to identify fishing sets and prospecting phases off which it was relevant to distinguish between cruising and tracking.

The state-space methodology has proven to be efficient to describe population dynamics (Ovaskainen and Hanski, 2001; Buckland et al., 2004; Thomas et al., 2005; Patterson et al., 2008), or to re-estimate the real trajectory of a tagged animal according to the associated state in a given landscape (Royer et al., 2005; Jonsen et al., 2005). This framework has also been applied to infer foraging and encamped phases in Canadian elks' movements (Morales et al., 2004). The primary objective of this work was thus to develop a state-space model to estimate the behavioural states of fishing vessels from their trajectories regularly sampled by VMS. Speeds and turning angles were extracted in sequences from vessels' trajectories, and were used to estimate fishing activity. Thanks to empirical evidences, we choose to model shifts between states by an order one Markovian process. This model was built in a Bayesian framework so that the inference of the model parameters was outputs of Monte Carlo Markov Chains (MCMC). To our knowledge no publication has yet been produced on the use of such model on VMS data even though this is an active field of research (Vermard et al., 2010). More importantly, none is validated by field observations. The secondary objective of this paper was thus to validate the estimations of such a Bayesian Hidden Markov model by real observations. Without such ad hoc validation, these studies remain speculative and we believe that the present paper represents an important step for the operational use of VMS in fisheries ecology.

Thanks to a long and fruitful collaboration with the French tropical tuna purse-seiners association and the French governmental department of Fisheries, we had access to their VMS data though under some confidentiality conditions. Meanwhile, due to the European Data Collection Regulation (DCR), 10% of the trips get observers on board providing us with data to validate the model.

2. Materials

2.1. Data

Since 2000, the European Commission legislated that all European fishing vessels longer than 24 m should be equipped with a Vessel Monitoring System (VMS) (and then all vessels longer than 15 m, in 2005) (EC, 1997 and 2003). The data used for this study were collected from the French-flagged purse-seiners based in the Seychelles islands, fishing several tropical tuna species (*Thunnus albacares*, *Katsuwonus pelamis*, *Thunnus obesus*) in the Western Indian Ocean (15 vessels in 2006, 17 vessels in 2007). The Global Positioning System (GPS) positions of the vessels were registered every hour and transmitted on shore by satellite (Argos or Inmarsat). Being GPS positions, the data were accurate (error smaller than few tens of meters) and regularly recorded every hour. Speeds (in knot) and turning angles (in rad) between consecutive positions were readily calculated from VMS data. They correspond to apparent hourly speeds and angles and not to the integration of the real speeds and angles along the vessel track. To honour the fact that vessels are most of time either full speed (around 12 knots) or immobile, we divided all speeds by the maximum possible speed for each vessel (~12 knots) so that they ranged between 0 and 1. Given the shoaling behaviour of the tropical tuna and given the fact that fishermen based their decision to fish on visual detection of tuna schools, fishing activity occurs at day. Only the daytime parts of the trajectories were then used. Starting and ending time of the day were deduced from the date and the latitude and longitude of each GPS position through an ad hoc routine to automatically select the daylight VMS data.

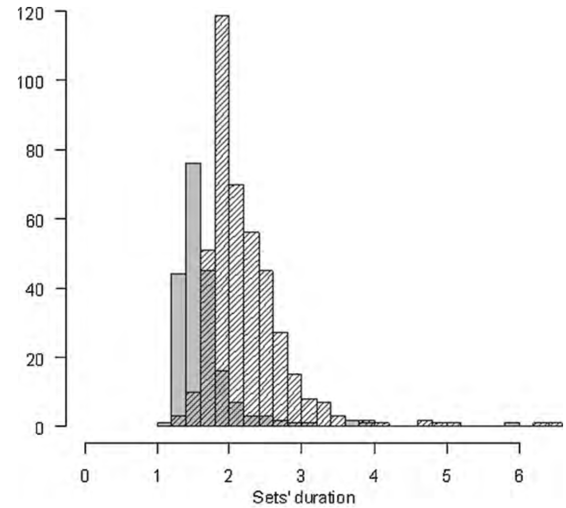

Fig. 1. Histograms of the duration (in hours) of fishing operations (fishing sets). Observers' data set (10% of the complete fleet). Hachured white: successful sets ($m = 2.23$ h, $\sigma = 0.58$ h). Grey: unsuccessful sets ($m = 1.65$ h, $\sigma = 0.50$ h).

The observers' program in the French purse-seiner fleet in the Indian Ocean is being undertaken in the framework of the European Data Collection Regulation (DCR – EC Reg. 1563/2000). This regulation specifies that 10% of the trips realised by each member country fleet have to get an observer on board. For the French fleet of tropical purse-seiners operating in the Indian Ocean, the program started in November 2005 and 16 trips over a total of 120 trips were available for the period 2006–2007. However, only 11 trips corresponded to vessels with a French flag reducing the number trips that were available for the analysis. Observers record the position of the vessel at least every hour, and at each change of speed or turning angle (course) of the vessel. They also record the beginning and ending time of each fishing operation. One of these trips considered as a standard one with regards to the number of days at sea, the number of fishing sets, and the ratio of sets on free-swimming schools and log schools, was used for calibrating the parameters of the model.

2.2. States and activities

We distinguished the states of a vessel (movement states which refer to its trajectory) from its activity. Schematically, the primer concerns the vessel, while the latter refers to the fishermen on board of this vessel. For the tropical tuna purse-seiner studied here, four major activities were considered. A fishing operation (activity 1 = fishing) lasted generally more than 1 h due to the time required to set the seine out and to get it back. Unsuccessful fishing sets (null sets) corresponded to stops of 1.65 h on average ($\sigma = 0.5$ h; Fig. 1). When the school was effectively trapped (positive sets), the extra time required to brail fish on board was 1 h per 100 tonnes on average. The average duration of a positive fishing operation was then 2.23 h ($\sigma = 0.58$ h; Fig. 1). The second type of activity (activity 2 = stop) corresponded to stops at sea. Stops were required either to maintain electronic equipments located on Fish Aggregating Devices (FADs) or to evaluate the effective presence of fishable schools. In this regards stops at sea contributed, even in a non-trivial manner, to the fishing effort. Long stops were hardly ever due to technical break-downs and damages. As a matter of fact, when such problems occur, skippers try as much as possible to postpone the immobilization of the vessel until the night. It was thus hypothesised that daylight stops corresponded to real stops contributing to the fishing activity. In the case of purse-seiner, no movement of the vessels (state 1 = still) was thus associated to two different activities namely fishing or stop at sea (state 1 = activity 1 + activity 2). The third activity was defined as the active tracking of visual appearances of tuna at the surface (birds, “balbaya”, “sardara”, FAD). This activity was typically associated to tortuous trajectories which was associated to the second key possible movement state of the vessels (activity 3 = state 2 = tracking). The final activity considered was cruising. In the particular case of purse-seiners, fishermen systematically maintain radar and visual tuna school inspection during all daily cruising phases (even when going straightforward to a new fishing site). Cruising phases are thus clearly part of the fishing effort even if trajectories associated to them were linear (activity 4 = state 3 = cruising).

We got then three behavioural states and four fishing activities to handle (Table 1). Only one activity corresponded to fishing, strictly speaking. This was also the only activity recorded by observers. The other three ones were, in more or less complex manners, connected to prospecting effort and were not recorded by observers. Calibration of the model parameters and validation of the model outputs were thus expressed in terms of “fishing” versus “non-fishing” activities (Table 1).

2.3. Moves and steps

One usually distinguishes two different frameworks for analysing trajectories (Turchin, 1998, pp. 127–134). In the step by step approach, trajectories are regularly sampled at a resolution given by the size of the steps in time. The true track between two samples (recording positions) is not known and the distributions of apparent speeds and turning angles provided by the observations are different from the true ones and vary with the step value (the homogeneity of the time resolution for the steps is thus fundamental). Alternatively, one can define the parts of the trajectory with constant speed and direction (moves). Moves are defined by constant speed and direction between two breaking points. The tracks between two breaking points are known and linear (up to some unavoidable tolerances needed in practice).

VMS positions were recorded every hour without any missing position and belonged to the primary type of data (steps). They were available for all trips (exhaustiveness over the fleet). Observers’ data, and in particular their disjunctive “fishing/non-fishing” information, belonged to the second one (moves). With the objective to use them to improve the inference of vessels’ activity from VMS data (i.e. definition of informative priors, definition of an optimal threshold to discriminate fishing from stopping, evaluation of the quality of the estimations), each step of trips with an observer was assigned a fishing or a non-fishing activity (Fig. 2). Due to the different temporal supports of the step data (fixed and equal to 1 h), and of the move data (variable), and given that fishing was the key targeted activity, each step in which fishing occurred at least thirty minutes was considered to be a fishing step. Given the histogram of the duration of the fishing sets (Fig. 1), a fishing operation covered between two and seven consecutive steps.

3. Methods

3.1. Dominant state and activity

The objective was to infer the activity of the vessel within each step knowing the apparent speeds and turning angles per step and to deduce which parts of the whole trajectory corresponded to fishing operations. In essence, states and activities are instantaneous qualitative variables. Their definitions at the step support were based on the following considerations. First, vessels usually did not switch from one particular state to the others too often, so that most of the steps often coincided with one unique state (pure steps). Second, changes of activity (moves) did not coincide with VMS acquisition (steps) so that some steps were made of several activities. Third, in 1 h-period of time, fishermen could not experience more than two different types of fishing dynamics so that steps were made of two different activities at most, of which one is necessarily dominant and considered as the activity for the step.

3.2. Model

Vessels’ activities were estimated in two phases. In the first phase, a model of order 1 Markovian hidden states was used. A Bayesian framework allowed getting posterior distributions for the possible states of all steps. The second phase aimed at disentangling still state (state 1) between fishing and stops (activities 1 and 2). Finally, validation ought to quantify the ability to detect properly the presence/absence of fishing sets.

3.2.1. The state-space model

The unknown states of the vessels were represented by an hidden (latent) random multinomial variable denoted $I_t \in \{1,2,3\}$, for each step t. Observations (y_t) corresponded to pairs of apparent relative speeds (s_t) and turning angles (ϕ_t) of each step, and were considered as outcomes of the random variables S_t and Φ_t. Relative speeds were considered to be *Beta* distributed with parameters depending on the state:

$$[S_t|I_t = i] = Beta(a_i, b_i), \quad \text{for } i \in \{1, 2, 3\} \tag{1}$$

with *Beta* probability density function defined by:

$$f_{Beta}(s_t, a_i, b_i) = \frac{1}{B(a_i, b_i)} \cdot {s_t}^{a_i-1} \cdot (1 - s_t)^{b_i-1} \tag{2}$$

where $a_i,b_i > 0$ and $B(a_i,b_i) = \Gamma(a_i)\cdot\Gamma(b_i)/\Gamma(a_i + b_i)$.

Similarly, turning angles were modelled by a Wrapped Cauchy distribution (Bartumeus et al., 2008) with parameters depending on the state:

$$[\Phi_t|I_t = i] = WC(\mu_i, \rho_i), \quad \text{for } i \in \{1, 2, 3\} \tag{3}$$

The Wrapped Cauchy distribution is a circular probability function that allows modelling angular variables. It belongs to the same family of laws than the Von Mises distribution or the Wrapped Normal distribution (Mardia and Jupp, 2000; Jammalamadaka and SenGupta, 2001). Its probability density function is:

$$f_{WC}(\phi_t, \mu_i, \rho_i) = \frac{1}{2\pi}\cdot\frac{1 - \rho_i^2}{1 + \rho_i^2 - 2\rho_i \cdot \cos(\phi_t - \mu_i)} \tag{4}$$

3.2.2. Markovian property

States were unknown, making hazardous any decision on their properties, and in particular on their temporal autocorrelation. However, speeds and turning angles being dependent on the states, their statistical properties were used to infer whether the states could be modelled by a Markov process and, if yes, of which order. Coefficients of correlation between successive speeds were large ($\rho(S_t,S_{t+1}) = 0.58$, $P < 0.001$). The partial correlations, which quantify the remaining correlation between the speeds at step t and $t + 2$, after the correlation with the intermediate speed is removed, were much lower ($\rho(S_t,S_{t+2}|S_{t+1}) = 0.06$). This indicated that the speed at a given step t was largely dependent of the previous speed but rather independent from the speed two steps before. The distribution of the turning angles also showed a mode around 0 (Fig. 3) not consistent with a random orientation at each step (Bovet and Benhamou,

Table 1
Definition of the dominant states and activities per step. Discrimination between effective fishing and various activities contributing to prospection.

Model (120 trips, 96,325 steps)		Observations (11 trips, 3510 steps)
States	Activities	
Cruising	Cruising	Non-fishing
Tracking	Tracking	(prospecting activities;
Still	Stop	components of fishing effort)
	Fishing	Fishing

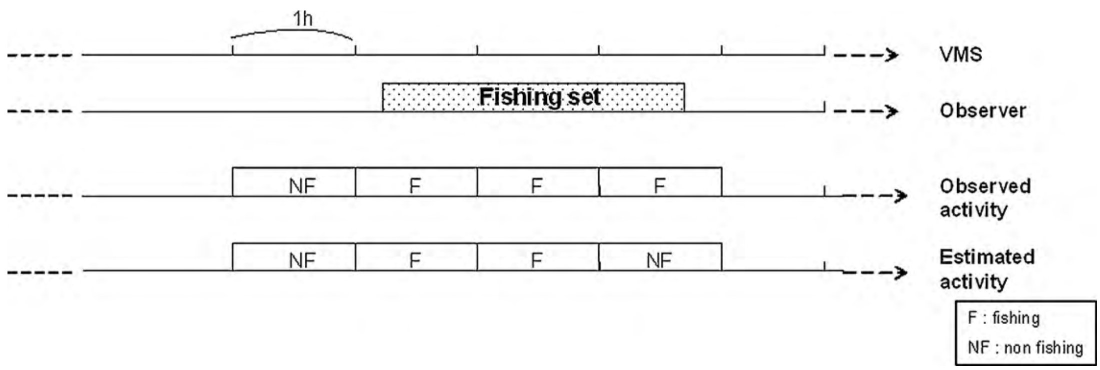

Fig. 2. Rule for affecting "fishing"/"non-fishing" activities into regular steps. This was compulsory to confront model outputs (steps) to observations (sets). A step was said to be "fishing" (reversely "non-fishing") when at least 30 min of fishing occurred during the step (dominant activity).

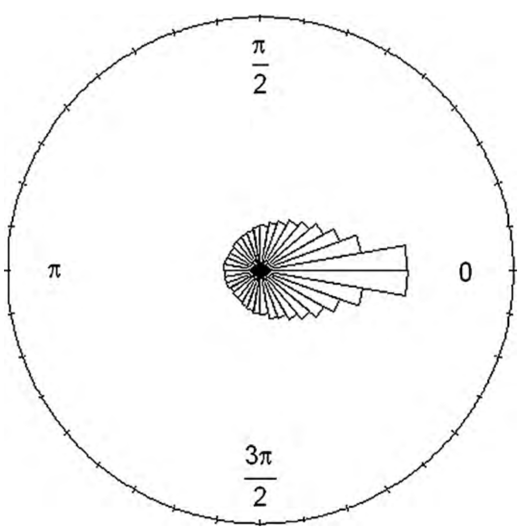

Fig. 3. Histogram of turning angles (in rad) for all the trajectories (96,325 steps).

1988; Turchin, 1998). Based on these empirical evidences, a Markov property of order 1 was then chosen for the states.

The associated transition matrix was designed considering that vessels could not switch directly from still state to cruising state, and vice versa (probabilities equal to 0):

$$M = \begin{array}{cl} \begin{bmatrix} p_1 & 1-p_1 & 0 \\ p_2 & 1-p_2-p_3 & p_3 \\ 0 & p_4 & 1-p_4 \end{bmatrix} & \begin{array}{l} \text{still} \\ \text{tracking} \\ \text{cruising} \end{array} \\ \begin{array}{ccc} \text{still} & \text{tracking} & \text{cruising} \end{array} & \end{array} \tag{5}$$

The model was thus based on six parameters for the *Beta* distributions of the speeds (three states times two parameters per state), six parameters for the Wrapped Cauchy distributions of the turning angles, and four parameters for the transition matrix. All these 16 parameters were denoted $\theta = (a_1, a_2, a_3, b_1, b_2, b_3, \mu_1, \mu_2, \mu_3, \rho_1, \rho_2, \rho_3, p_1, p_2, p_3, p_4)$.

3.3. Bayesian inference of the states

3.3.1. Simplification of the general expression

In a Bayesian framework (e.g. Strenio et al., 1983; Clayton and Kaldor, 1987; Wikle, 2003; Gelman et al., 2004; Morales et al., 2004), the (posterior) probability of the state values and of the parameters was proportional to the likelihood of the data times some prior distribution for the parameters

$$[I_{1\ldots T}, \theta | y_{1\ldots T}] \propto [y_{1\ldots T} | I_{1\ldots T}, \theta] \cdot [I_{1\ldots T} | \theta] \cdot [\theta] \tag{6}$$

Given the assumptions above mentioned, the posterior distribution simplified itself into

$$[I_{1\ldots T}, \theta | y_{1\ldots T}] \propto \left(\prod_{t=1}^{T} [y_t | I_t, \theta] \right) \cdot \left(\prod_{t=2}^{T} [I_t | I_{t-1}, \theta] \right) \cdot [I_1] \cdot [\theta] \tag{7}$$

Assuming that the speeds and turning angles were (conditionally) independent, the likelihood function ultimately reduced to the following expressions:

$$\prod_{t=1}^{T} [y_t | I_t = i, \theta] = \prod_{t=1}^{T} [S_t, \Phi_t | I_t = i, \theta] = \prod_{t=1}^{T} f_{Beta}(s_t, a_i, b_i).f_{WC}(\phi_t, \mu_i, \rho_i)$$
$$\text{for } i \in \{1, 2, 3\} \tag{8}$$

3.3.2. Definition of informative priors

One of the 11 trips with observers on board, i.e. a set of data for which the states were known, has been sacrificed to derive posterior distributions for the parameters. These posterior distributions were further used to define informative priors for the rest of the procedure.

As observers only recorded fishing versus non-fishing activities, states were temporally grouped to define the following new binomial variable:

$$I'_t = \begin{cases} 1 & \text{if} \quad I_t = 1 \\ 0 & \text{if} \quad I_t = 2 \quad \text{or} \quad I_t = 3 \end{cases} \tag{9}$$

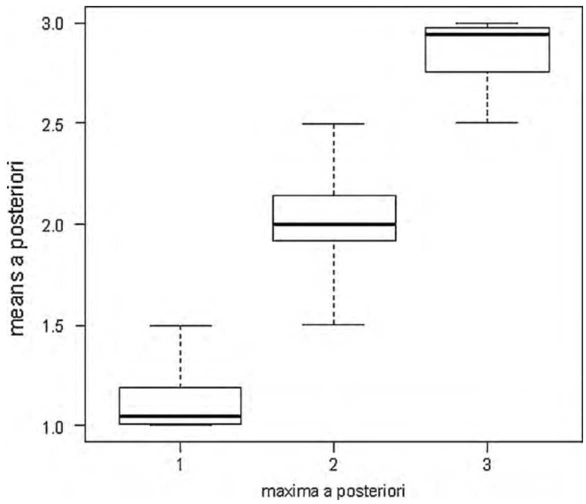

Fig. 4. Relationships between maxima a posteriori (MAP) and means a posteriori (mAP). For each of the 96,325 steps, the posterior multinomial distribution of the possible states allowed computing a MAP (the state getting the largest probabilities to occur) and a mAP (the arithmetic means of the posterior distributions). For each possible MAP value (*x*-axis), the boxes represent the 25% and 75% quartiles of the 96,325 associated mAP values, and the whiskers represent the extreme mAP values.

The posterior distribution for the parameters was thus:

$$[\theta|I'_{1\ldots T}, y_{1\ldots T}] \propto [y_{1\ldots t}|I'_{1\ldots T}, \theta] \cdot [\theta|I'_{1\ldots T}] \quad (10)$$

We set vague priors (uniform distributions) for all parameters and got posterior distributions that were (visually) Gaussian.

3.3.3. *Prediction of states with informative priors*

In order to get a workable model, we fixed above defined informative priors for the parameters of the probability density function of the speeds (a_1,a_2,a_3,b_1,b_2,b_3) and of the turning angles (μ_1,μ_2,μ_3,ρ_1,ρ_2,ρ_3). To honour possible differences from one vessel to the others, we let non-informative priors for the parameters of the transition matrix.

Posterior distributions for the states were obtained by MCMC simulations using the software OpenBUGS (Thomas et al., 2006) in the software R with the package BRugs (R Development Core Team, 2004). The convergence was tested by the Gelman–Rubin test (Gelman and Rubin, 1992; Brooks and Gelman, 1998) with three chains of 20,000 iterations each. The Gelman–Rubin statistic reached 1 after 2000 iterations. A burn-in period of 2000 iterations was thus used and each chain was run during 20,000 extra iterations to get posterior distributions.

The model was applied vessel by vessel allowing the parameters to differ from one vessel to the next. This was done to honour the fact that different vessels have different fishing strategies, that is different manners to switch from one activity to another and different resilience to activities. These differences translated potentially into different transition matrices. It was however assumed that, for a given activity, the distributions of the speeds and turning angles were similar over the entire fleet.

3.3.4. *MAP and mAP*

For each step, we extracted both the maximum a posteriori (MAP) and the mean a posteriori (mAP). Being multinomial, only the MAP for the states got a straightforward meaning. However, the mAPs were linked to the MAPs in such a way that the states' estimates could be interpreted as two pairs of binomial variables (Fig. 4). This was of particular interest as this allowed using the mAPs as an indicator for the uncertainties around the estimate.

3.4. *From states to activities*

The aim of the last phase of the procedure was to automatically identify which of the still steps were stops at sea or fishing sets. Given the characteristics of the fishing operations and of the frequency of data acquisition (see above), fishing sets spread over at least 2 consecutive steps. We thus considered that steps that belonged to sequences of at least two steps with mAP smaller than 2, preceded and followed by steps with mAP equal to 2 or larger than 2, could potentially be fishing steps. Amongst these pre-selected steps, steps below a threshold, yet to be defined, were finally selected.

The threshold was defined by an iterative procedure seeking the minimum overall misdetection rate between estimates and truth (observers' data). It was implemented on the 10 remaining trips with observers on board.

3.5. *Validation*

The best validation being the confrontation to field data, we confronted the estimated activities to those declared by the observers.

However, given the characteristics of the data this was not trivial.

The very truth data were provided by observers on board 10% of the vessels. Their declarations were based on fishing events only (not on states nor activities) and did not conform in terms of time of acquisition to VMS acquisition. A first calibration–validation phase corresponded to the allocation of still states to "fishing" or "stop" activities (Fig. 5). The calibration corresponded to the best correspondence between model outputs and observers' data at step level with regards to (dominant) fishing activity (Fig. 2). Given that one trip had been used to infer informative priors, it remained only 10 trips with full information for this calibration–validation phase and we decided to use them all to estimate the optimal cut-off, i.e. value for which the error rate between estimated fishing and observed fishing was minimum.

This was preferred than splitting the 10 trips with observers into a calibrating and validating data sub-sets as we would have been unable to interpret the unavoidable difference of the model performance in the two sub-sets and as we would have lost a large amount of data for each of the two phases (calibration and validation).

The code used for the analyses is attached in Annex 1.

4. Results

4.1. *Speeds and turning angles per state*

The mAP for the parameters allowed getting the distributions of speeds and turning angles for the three states (Fig. 6). For the "still" state, the distribution of speeds was characterised by a mode at zero, and the distribution of turning angles was uniform. These distributions corroborate the behaviour of the vessel during a stop near a free-swimming school or a school associated with a FAD, or during a fishing set (Fig. 6a and d).

On the contrary, speeds and turning angles during "cruising" had sharp modes around 12 knots and 0 rad (Fig. 6c and f). This was strongly consistent with the fact that when they were cruising, vessels went straight at full speed, in order to move to another area or to go to a FAD (equipped with a proper radio buoy) without changing their direction (and speed) until appearance of tuna school was sighted.

The "tracking" state depicted intermediate distributions however more influenced by the "cruising" state. The distribution of speeds was more tailed with a smooth mode around 10 knots and the turning angles kept a mode around 0 rad. During the active

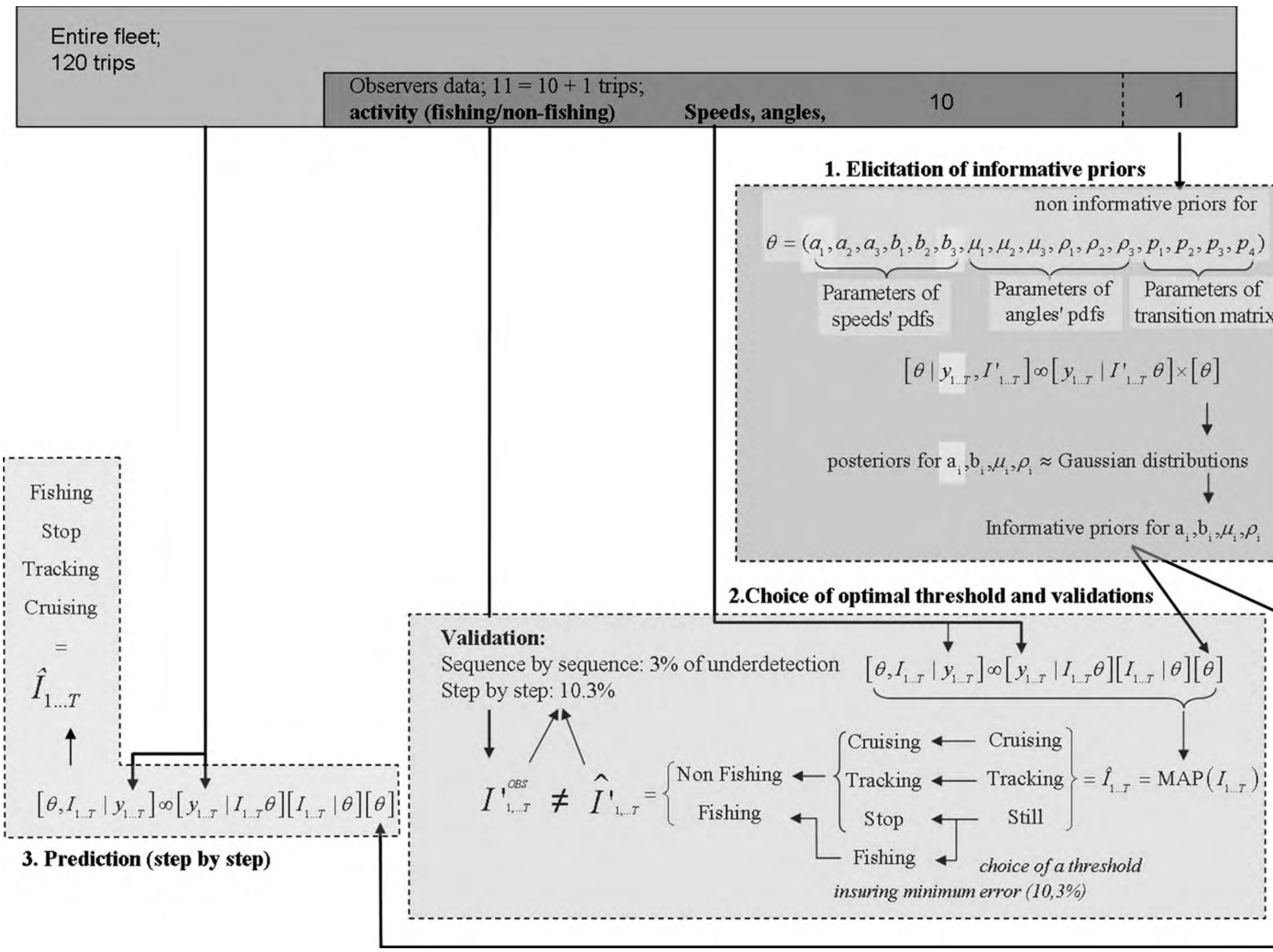

Fig. 5. Flowchart of the complete methodological framework. The main phases are (1) the use of one of the eleven trips with full information for the elicitation of informative priors, (2) the use of the 10 remaining trips to choose a threshold to disentangle still states into effective fishing operations and stops at sea and to quantify the error made at this stage, and the validation of fishing steps with the 10 remaining trips, and (3) the use of all vessels' locations (Vessel Monitoring System data) to estimate vessels' dominant activities during each step.

search of tuna schools in an area of tuna aggregation, the changes in course were highly frequent and the vessels were cruising at full speed. But with hourly data, the apparent mean speed is necessarily intermediate (10 knots) (Fig. 6b and e).

4.2. Quality of the estimations

The use of observers' data to tune the optimal threshold to discriminate between "fishing" and "stop at sea" led to a compromise between under-detections and over-detections of fishing steps. The overall error did not show a clear optimal between threshold values of 1.1 and 1.6 (Fig. 7). The lowest mean square error (10.2%) was obtained for a threshold of 1.3. Despite the fact that similar level of under-detections and over-detections were obtained for a smaller threshold (around 5% for 1.2) without modifying significantly the overall MSE, we stuck to 1.3 which was more conservative with regards to under-detections (we did not want to miss fishing steps as much as possible without spoiling the overall performance of the model). This misdetection level was computed at the step level and was not relevant with regards to the capacity of the procedure to properly identify fishing operations as a whole. As a matter of fact, for say, a set lasting 2 and half hours (Fig. 2), the last step may be estimated to be "fishing" while this was not the case (in terms of dominant activity).

4.3. Fishing behaviour and effort

For each vessel, we took the mean and standard deviation of the posterior distributions of the parameters of their respective transition matrix. Differences between vessels were small (Fig. 8). The three probabilities on the diagonal of the transition matrix which corresponded to the probability to stay in the same state when moving to the next step, were homogeneously large amongst the fleet. The probabilities for changing from one state to the other (p_2, p_3 and p_4) were, on the contrary, homogeneously low over the fleet. There was a clear linear link between posterior means and standard deviations (Fig. 9) so that the coefficients of variation for the estimations of these three parameters were highly constant and low (10%).

The posterior means of the states were distributed over three clear modes around 1, 2 and 3 (Fig. 4). The frequencies of (the time spent in) each activity indicated that the fleet spent 78% of its time prospecting (i.e. daytime minus the fishing operations which represented though 22%). The prospecting phase broke down into 34% cruising, 61% tracking schools in areas of aggregations in sinuous trajectories, and 5% being immobile. The model outputs for one particular trip were geo-referenced (Fig. 10) so that the fishing effort could be analysed spatially.

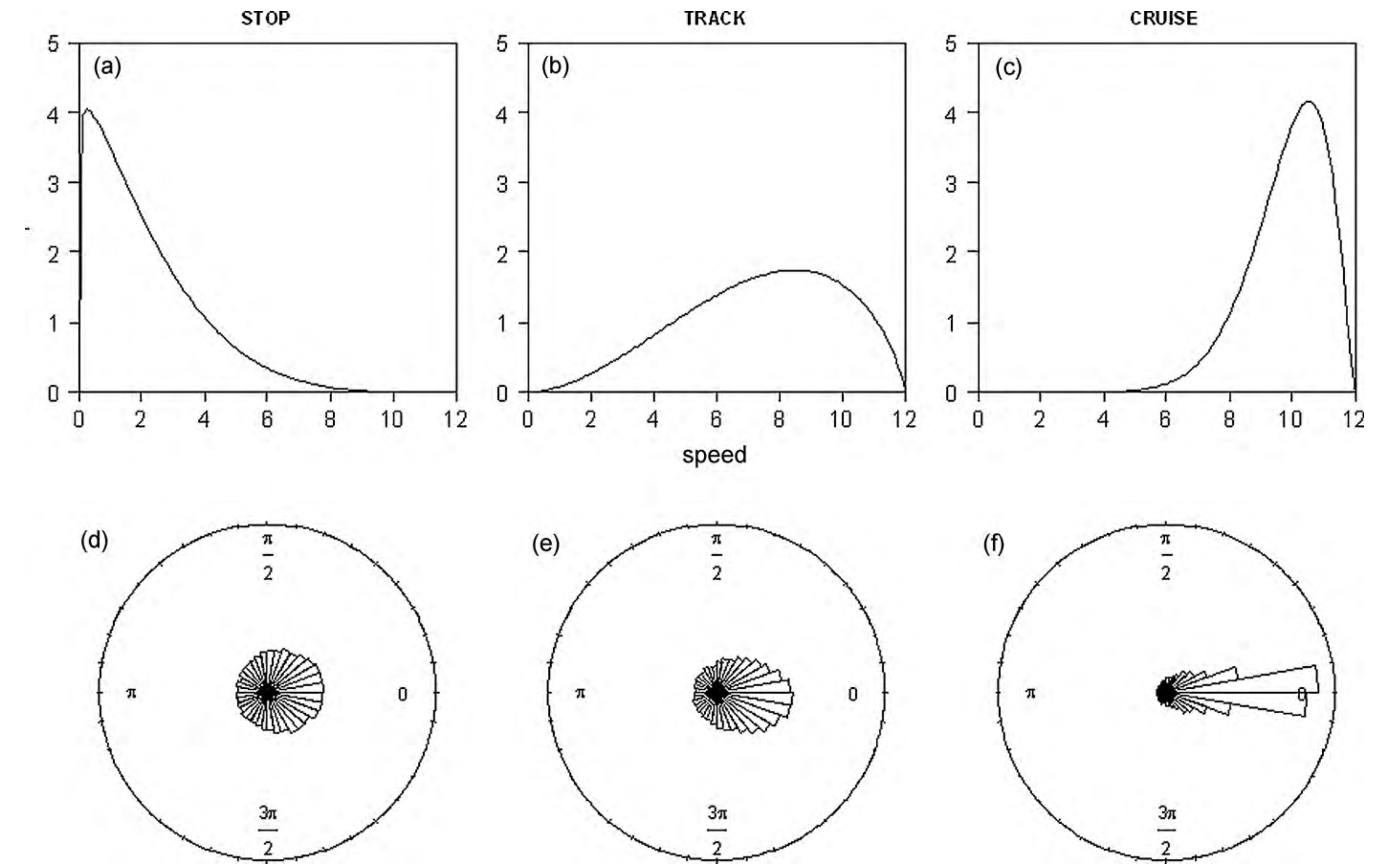

Fig. 6. Distributions of speeds (in knots, i.e. nautical miles per hour) (a–c) and turning angles (in rad) (d–f) for the three states.

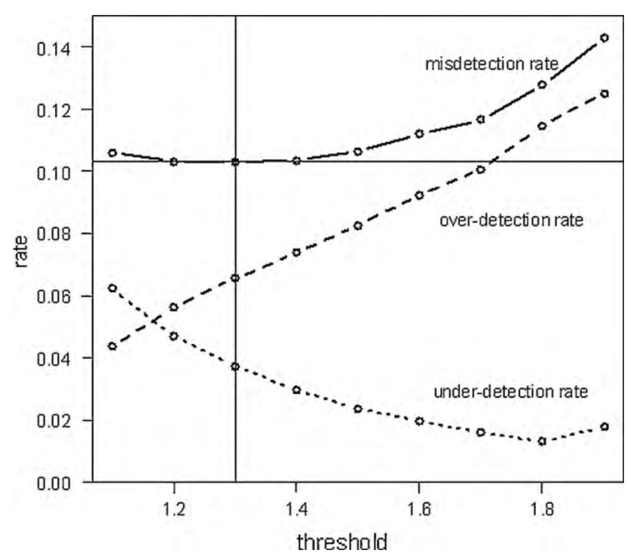

Fig. 7. Determination of an optimal mAP threshold for affecting one of the two possible activities "stop" and "fishing" to steps estimated in "still" state. Errors are evaluated as the proportion of correctly classified steps with regards to "fishing"/"non-fishing" activities by confronting model outputs and observers' declarations (Table 1) using the 10 available trips with both VMS and observers' data (3510 steps). The total proportions of misdetection (*y*-axis) are decomposed into over-detection rates and under-detection rates as a function of the threshold values (*x*-axis). The vertical line corresponds to the chosen threshold (1.3).

5. Discussion

The Bayesian Hidden Markov model developed in the present study was quite similar to already published works (Morales et al., 2004; Jonsen et al., 2005; Vermard et al., 2010). The model was indeed a mixture of the "triple switch" model (three states with switching probabilities between states) and of the "switch constrained" model developed by Morales et al. (2004). The main difference between our work and other analyses of animal trajectories (Morales et al., 2004; Jonsen et al., 2005) and of VMS trajectories (Mills et al., 2007; Witt and Godley, 2007) relied on the use of field data provided by observers on board 10% of the vessels. This allowed (i) setting specific values for the probability density functions of the speeds, (ii) setting informative priors for the parameters of the transition matrix, (iii) choosing the threshold to disentangle fishing events from stops at sea and (iv) qualifying the output by a rigorous validation phase. In a matter of fact, most analyses of trajectories, if validated, are validated by qualitative means. Amongst these qualitative criteria, one can mentioned the coherence of the distributions of speeds and turning angles by state (e.g. Morales et al., 2004; Vermard et al., 2010), the sinuosity of tracks in ARS areas (e.g. Tremblay et al., 2007) or simply the fact that algorithms managed to provide outputs (Jonsen et al., 2005). The ARS areas estimated by an algorithm based on fractal dimension (Tremblay et al., 2009) were consistent with the tracking steps estimated by the actual model. However, such algorithms are specifically dedicated to the detection of sinuous patterns and can identify neither fishing sets nor cruising path. The same type of approach was already applied to animals' trajectories (Morales et al., 2004; Jonsen et al., 2005, 2007; Bailey et al., 2008), although no validation was performed. Some works in theoretical ecology showed that the characteristics of predators' trajectories linked to their searching strategy can be modelled by a correlated random walk (CRW) (Bovet and Benhamou, 1988; Zollner and Lima, 1999). Given that fishers' behaviour is approaching top predators' behaviour (Bertrand et al., 2007), the assumptions and validation

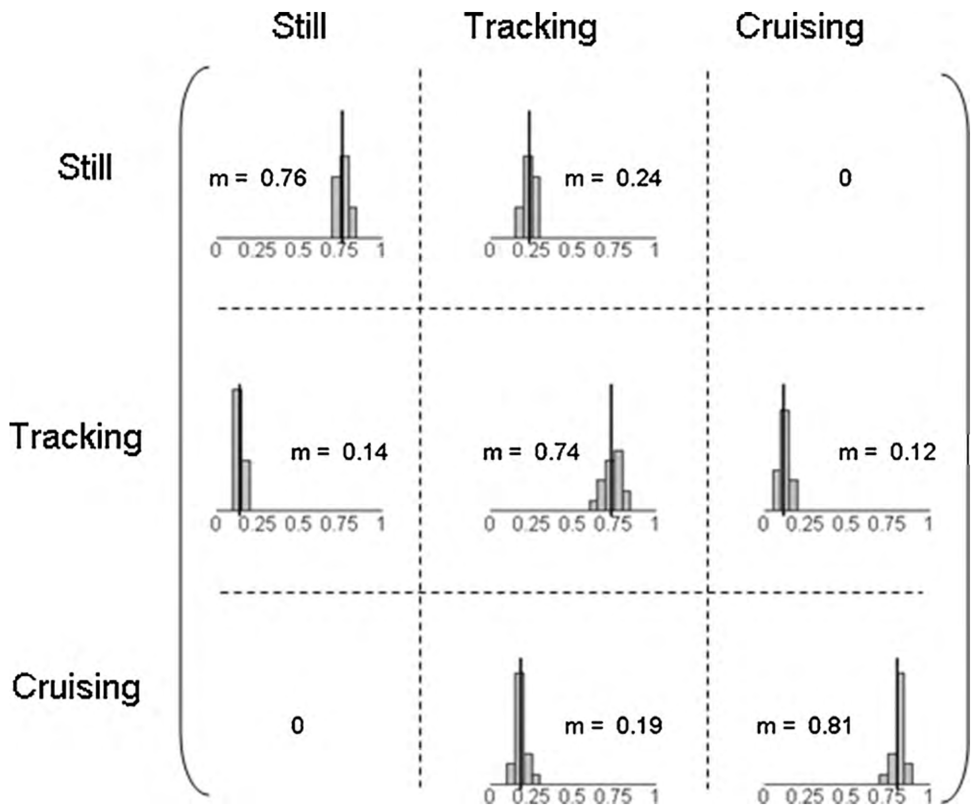

Fig. 8. Histograms of the 17 means of the posterior distributions of the parameters of the transition matrix of the 17 vessels available for this study. The overall mean is indicated and represented by a vertical line as well.

components of the present modelling may be adapted to other predators' trajectories.

The Markov property of state-space models is often postulated while in the present case we based this decision on empirical evidences. By the way, the model could have easily been adapted to semi-hourly VMS data (not presented here), but would have required fundamental changes in the case of smaller time intervals between positions (10 min) as the Markov property of order one would have became largely insufficient or in the case of irregular acquisition of positions. Similarly, the conditional independence between speeds and turning angle was an acceptable assumption because we used apparent speeds and angles for 1 h-steps, and not instantaneous values. For high resolution data, it would have been irrelevant to imagine a vessel making a 180° turn while being full speed (negative correlation between instantaneous speed and angle in state "cruising"). However, at the scale of 1 h-steps, apparent speeds and angles can be considered as having no correlation (knowing the speed does not tell any thing on the turning angle provided that we know the state). The fact that both the probability density functions of the speeds and of the turning angles changed together with the state, is indeed a consequence of the conditional independence present in the model.

Observers' data allowed both choosing the structure of the model (e.g. Markov behaviour), setting informative priors for some parameters and quantifying the performance of the model. We assumed (i) that fishermen's behaviours were different in terms of their resilience in a particular activity or their tendencies to move quickly from one activity to the other and (ii) that the speeds and turning angles were specific to an activity and not to a fishermen. We thus let the transition matrix to move from one vessel to the next (but the outputs indicated that little differences existed) and let distributions of speeds and turning angles common to all vessels. As far as priors are concerned, we distinguished between parameters that were assumed independent from the vessel and those possibly dependent. The first sub-set corresponded to the parameters of the speeds and turning angles probability density functions. The second corresponded to the transition matrix. The use of informative versus non-informative priors was however tested and did not show any significant impact on the results. Still the use of informative priors allowed to speed up the convergence of the Monte Carlo Markov Chains, and to insure coherence in the sequence of activities with regards to postulated links between tuna presence and activity (not presented here).

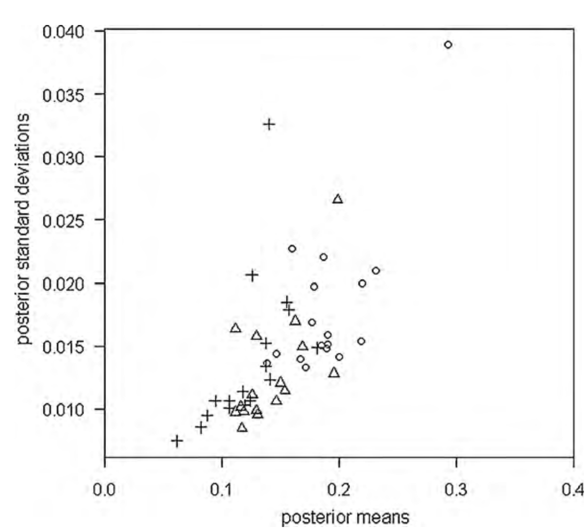

Fig. 9. Relationship between posterior means and standard deviation for the parameters of the transition matrix of the 17 vessels. Each point corresponds to the mean and standard deviation of the posterior distributions for one parameter of the transition matrix. Triangles: p_2; crosses: p_3; circles: p_4 (see Eq. (5)).

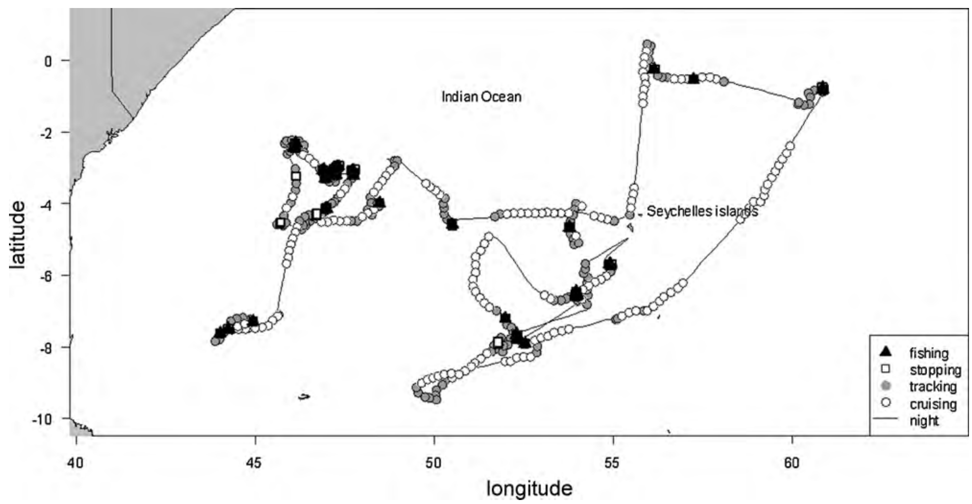

Fig. 10. Outputs of the model ("fishing", "stop", "tracking", and "cruising") are represented geographically along the trajectory of a vessel (example of one randomly selected trip of 30 days). Night parts of the track are represented by thin continuous lines.

However, using observers' data had a price. We had to make compromises when crossing observations and trajectories, specifically with regards to fishing operations. The main difficulty was in the fact that moves data (observers) did not have the same time resolution than step data (VMS). We solved the problem by defining dominant activities which remain an operational solution as long as the frequency with which skippers move from one activity to another is not too high (high with respect to the frequency of VMS acquisition). Still there were many steps that corresponded to several activities. The question that arose was then to know if the mAP of state could be interpreted as estimation of the proportion of time spent by the vessel in different states. The mAP is clearly an empirical version of the expect value of the multinomial state random variable and is, in this regard, clearly defined. However, its meaning in terms of proportion of states is unclear and non univocal. For instance, in theory, a mean of 2.5 can be either associated to equi-probable occurrence of two and three or, with three quarters of threes and one quarter of ones. The above question was indeed answered in two steps. First the relationship between MAP and mAP (Fig. 4) indicated that despite the fact that the state was in theory a multinomial variable, in practice, it could be reduced to sets of binomials for the interpretation. In such a case, the mAP became directly connected to the probabilities associated to the two competing states. It was then further assumed that the hesitation to elucidate a dominant state for the running step, and though the value of the mAP, was an estimation of the proportion of the time spent in each states. In other words, intermediate mAP values (between 1 and 2, and between 2 and 3) indicated cases where one part of the hourly step corresponded to one type of activity, and the second to an other one.

Because vessels were, up to a small drift, totally immobile during fishing operation, and because fishing sets lasted at least 1 h 15 min, a simple discrimination between the low hourly speeds and the high hourly speeds to detect fishing operations were used as a benchmark method (i.e. a level above which any more sophisticated method should not go). The optimal threshold we obtained provided 20% of misdetections. Surprisingly, this is of the same order of magnitude as an even more simple method where all steps were classified as non-fishing. The proportion of the fishing steps in the observers' dataset being 19%, a decision rule attributing systematically a non-fishing state to all steps would also generate 19% of misclassifications. The method presented in this paper provided twice better outputs than the benchmark models (10.2% misclassified steps). By the way, 97% of the fishing sets declared by observers had been detected by the model. Some sequences (e.g. three fishing sets of 2 h each separated by only a quarter of an hour) were viewed as only one long sequence of six fishing steps which explains possible small under-detection. The percentage of over-detection (15%) appeared large. Two elements can be advocated here. First, compared to the performance of the model at step level, this percentage was computed over the number of fishing operations which was two orders of magnitude lower than the number of steps. Second, given that in this particular fishery, fishing is equivalent to stillness and that vessels have sometimes to stop, it is impossible to avoid interpreting stop sequences as fishing sequences.

The value of distinguishing four activities, and not only "fishing" or "not fishing", is to identify the change points of strategy, that is the skipper's decision to search actively (rigorous prospecting) tuna in a small area, or to move to another area (Gaertner and Dreyfus-Leon, 2004). Furthermore the stop and tracking states are proxies for tuna's presence and will allow quantifying the fishing effort corresponding to the fishing sets in the same area and at the same period. The fishing effort will have to be calculated for each cell of space and for a defined temporal scale. As it has been studied by Piet and Quirijns (2009), the scale of effort calculation has to be chosen wisely according to the fishery. The searching behaviour being a proxy for tuna presence, maps of potential presence of tuna could well be derived from the model's results helping the delimitation, of tuna concentrations in time and space.

Acknowledgements

The work was co-funded by the European project: "Catch, Effort and Discard Estimates in Real time" (CEDER, FP7-022615) and by CLS (Collecte Localisation Satellite in Toulouse, France, http://www.cls.fr). Thanks to cooperative skippers from ORTHONGEL, data from scientific observers were kindly provided by the French "Observatoire thonier tropical" (www.ot.ird.fr) from IRD. VMS data were provided by the French Fisheries Department (Direction des Pêches Maritimes).

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.ecolmodel.2010.05.007.

References

Bailey, H., Shillinger, G., Palacios, D., Bograd, S., Spotila, J., Paladino, F., Block, B., 2008. Identifying and comparing phases of movement by leatherback turtles using state-space models. J. Exp. Mar. Biol. Ecol. 356, 128–135.

Bartumeus, F., Catalan, J., Viswanathan, G.M., Raposo, E.P., da Luz, M.G.E., 2008. The influence of turning angles on the success of non-orientated animal searches. J. Theor. Biol. 252, 43–55.

Bertrand, S., Burgos, J., Gerlotto, F., Atiquipa, J., 2005. Lévy trajectories of Peruvian purse-seiners as an indicator of the spatial distribution of anchovy (*Engraulis ringens*). ICES J. Mar. Sci. 62, 447–482.

Bertrand, S., Bertrand, A., Guevara-Carrasco, R., Gerlotto, F., 2007. Scale-invariant movements of fishermen: the same foraging strategy as natural predators. Ecol. Appl. 17 (2), 331–337.

Bertrand, S., Díaz, E., Lengaigne, M., 2008. Patterns in the spatial distribution of Peruvian anchovy (*Engraulis ringens*) revealed by spatially explicit fishing data. Prog. Oceanogr. 79, 379–389.

Bovet, P., Benhamou, S., 1988. Spatial analysis of animals' movements using a correlated random walk model. J. Theor. Biol. 131, 419–433.

Brooks, S.P., Gelman, A., 1998. General methods for monitoring convergence of iterative simulations. J. Comput. Graph. Stat. 7, 434–455.

Buckland, S.T., Newman, K.B., Thomas, L., Koesters, N.B., 2004. State-space models for the dynamics of wild animal populations. Ecol. Model. 171, 157–175.

Clayton, D.G., Kaldor, J.M., 1987. Empirical Bayes estimates of age-standardized relative risks for use in disease mapping. Biometrics 43, 671–681.

Deng, R., Dichmont, C., Milton, D., Haywood, M., Vance, D., Hall, N., Die, D., 2005. Can vessel monitoring system data also be used to study trawling intensity and population depletion? The example of Australia's northern prawn fishery. Can. J. Fish. Aquat. Sci. 62, 611–622.

Fauchald, P., Tveraa, T., 2003. Using first-passage time in the analysis of area-restricted search and habitat selection. Ecology 84, 282–288.

Gaertner, D., Pagavino, M., Marcano, J., 1999. Influence of fishers' behaviour on the catchability of surface tuna schools in the Venezuelan purse-seiner fishery in the Caribbean Sea. Can. J. Fish. Aquat. Sci. 56, 394–406.

Gaertner, D., Dreyfus-Leon, M., 2004. Analysis of non-linear relationships between catch per unit effort and abundance in a tuna purse seine fishery simulated with artificial neural networks. ICES J. Mar. Sci. 61, 812–820.

Gelman, A., Rubin, D.B., 1992. Inference from iterative simulation using multiple sequences. Stat. Sci. 7, 457–511.

Gelman, A., Carlin, J.B., Stern, H.S., Rubin, D.B., 2004. Bayesian Data Analysis, 2nd ed. Chapman & Hall/CRC, 698 pp.

Grémillet, D., Kuntz, G., Delbart, F., Mellet, M., Kato, A., Robin, J.P., Chaillon, P.E., Gendner, J.P., Lorentsen, S.H., Le Maho, Y., 2004. Linking the foraging performance of a marine predator with local prey abundance. Funct. Ecol. 18, 793–801.

Gimenez, O., Rossi, V., Choqueta, R., Dehais, C., Doris, B., Varella, H., Vila, J.-P., Pradel, R., 2007. State-space modelling of data on marked individuals. Ecol. Model. 206, 431–438.

Gutenkunst, R., Newlands, N., Lutcavage, M., Edelstein-Keshet, L., 2007. Inferring resource distributions from Atlantic bluefin tuna movements: an analysis based on net displacement and length of track. J. Theor. Biol. 245, 243–257.

Jammalamadaka, S.R., SenGupta, A., 2001. Topics in Circular Statistics. World Scientific, River Edge, NJ.

Jonsen, I.D., Flemming, J.M., Myers, R.A., 2005. Robust state-space modelling of animal movement data. Ecology 86, 2874–2880.

Jonsen, I.D., Myers, R.A., James, M.C., 2007. Identifying leatherback turtle foraging behaviour from satellite telemetry using a switching state-space model. Mar. Ecol. Prog. Ser. 337, 255–264.

Knoppien, P., Reddingius, J., 1985. Predators with two modes of searching: a mathematical model. J. Theor. Biol. 114, 273–301.

Mandelbrot, B.B., 1977. Fractals, Form, Chance and Dimension. W.H. Freeman and Co., San Francisco, 365 pp.

Mardia, K.V., Jupp, P., 2000. Directional Statistics, 2nd ed. John Wiley and Sons Ltd.

Mills, C.M., Townsend, S.E., Jennings, S., Eastwood, P., Houghton, C.A., 2007. Estimating high resolution trawl fishing effort from satellite-based vessel monitoring system data. ICES J. Mar. Sci. 64, 248–255.

Morales, J.M., Haydon, D.T., Frair, J., Holsinger, K.E., Fryxell, J.M., 2004. Extracting more out of relocation data: building movement models as mixtures of random walks. Ecology 85, 2436–2445.

Ovaskainen, O., Hanski, I., 2001. Spatially structured metapopulation models: global and local assessment of metapopulation capacity. Theor. Popul. Biol. 60, 281–304.

Ovaskainen, O., 2004. Habitat-specific movement parameters estimated using mark-recapture data and a diffusion model. Ecology 85 (1), 242–257.

Patterson, T.A., Thomas, L., Wilcox, C., Ovaskainen, O., Matthiopoulos, J., 2008. State-space models of individual animal movement. Trends Ecol. Evol. 23 (2), 87–94.

Piet, G.J., Quirijns, F.J., 2009. The importance of scale for fishing impact estimations. Can. J. Fish. Aquat. Sci. 66, 829–835.

R Development Core Team, 2004. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria, 3-900051-00-3 http://www.R-project.org/.

Rijnsdorp, A.D., Piet, G.J., Poos, J.J., 2001. Effort Allocation of the Dutch Beam Trawl Fleet in Response to a Temporarily Closed Area in the North Sea. ICES, Copenhagen (Denmark), CM 2001/N:01.

Robinson, P.W., Tremblay, Y., Crocker, D.E., Kappes, M.A., Kuhn, C.E., Shaffer, S.A., Simmons, S.E., Costa, D.P., 2007. A comparison of indirect measures of feeding behaviour based on ARGOS tracking data. Deep-Sea Res. Part II 54, 356–368.

Royer, F., Fromentin, J.M., Gaspar, P., 2005. A state-space model to derive bluefin tuna movement and habitat from archival tags. Oikos 109, 473–484.

Sims, D.W., Witt, M.J., Richardson, A.J., Southall, E.J., Metcalfe, J.D., 2006. Encounter success of free-ranging marine predator movements across a dynamic prey landscape. Proc. R. Soc B 273, 1195–1201.

Strenio, J.F., Weisberg, H.I., Bryk, A.S., 1983. Empirical Bayes estimation of individual growth-curve parameters and their relationship to covariates. Biometrics 39 (1), 71–86.

Thomas, L., Buckland, S.T., Newman, K.B., Harwood, J., 2005. A unified framework for modelling wildlife population dynamics. Aust. N. Z. J. Stat. 47 (1), 19–34.

Thomas, A., Hara, O.B., Ligges, U., Sturtz, S., 2006. Making BUGS open. R News 6, 12–17.

Tremblay, Y., Roberts, A.J., Costa, D.P., 2007. Fractal landscape method: an alternative approach to measuring area-restricted searching behavior. J. Exp. Biol. 210, 935–945.

Tremblay, Y., Robinson, P.W., Costa, D.P., 2009. A parsimonious approach to modeling animal movement data. PLoS ONE 4, 3.

Turchin, P., 1998. Quantitative analysis of movement. In: Measuring and Modeling Population Redistribution in Animals and Plants. Sinauer Associates, Inc, 396 pp.

Vermard, Y., Rivot, E., Mahévas, S., Marchal, P., Gascuel, D., 2010. Identifying fishing trip behaviour and estimating fishing effort from VMS data using Bayesian Hidden Markov Models. Ecological Modelling 221, 1757–1769.

Viswanathan, G.M., Afanasyev, V., Buldyrev, S.V., Murphy, E.J., Prince, P.A., Stanley, H.E., 1996. Lévy flight search patterns of Wandering Albatrosses. Nature 381, 413.

Wikle, C.K., 2003. Hierarchical Bayesian models for predicting the spread of ecological processes. Ecology 84 (6), 1382–1394.

Witt, M.J., Godley, B.J., 2007. A step towards seascape scale conservation: using Vessel Monitoring Systems (VMS) to map fishing activity. PLoS ONE 2, 10.

Zollner, P.A., Lima, S.L., 1999. Search strategies for landscape-level interpatch movements. Ecology 80 (3), 1019–1030.