Research Article **Environmetrics**

Received: 5 September 2013, Revised: 17 September 2014, Accepted: 9 October 2014, Published online in Wiley Online Library

(wileyonlinelibrary.com) DOI: 10.1002/env.2319

An autoregressive model to describe fishing vessel movement and activity

P. Gloaguen[a]*, S. Mahévas[a], E. Rivot[b], M. Woillez[c,b], J. Guitton[b], Y. Vermard[d] and M. P. Etienne[e]

The understanding of the dynamics of fishing vessels is of great interest to characterize the spatial distribution of the fishing effort and to define sustainable fishing strategies. It is also a prerequisite for anticipating changes in fishermen's activity in reaction to management rules, economic context, or evolution of exploited resources. Analyzing the trajectories of individual vessels offers promising perspectives to describe the activity during fishing trips. A hidden Markov model with two behavioral states (steaming and fishing) is developed to infer the sequence of non-observed fishing vessel behavior along the vessel trajectory based on Global Positioning System (GPS) records. Conditionally to the behavior, vessel velocity is modeled with an autoregressive process. The model parameters and the sequence of hidden behavioral states are estimated using an expectation–maximization algorithm, coupled with the Viterbi algorithm that captures the most credible joint sequence of hidden states. A simulation approach was performed to assess the influence of contrast between the model parameters and of the path length on the estimation performances. The model was then fitted to four original GPS tracks recorded with a time step of 15 min derived from volunteer fishing vessels operating in the Channel within the IFREMER RECOPESCA project. Results showed that the fishing activity performed influenced the estimates of the velocity process parameters. Results also suggested future inclusion of variables such as tide currents within the ecosystem approach of fisheries.

Keywords: hidden Markov model; vessels dynamics; RECOPESCA; autoregressive process; Baum–Welch algorithm

1. INTRODUCTION

Understanding the dynamics of fishing vessels is essential to characterize the spatial distribution of the fishing effort on a fine spatial scale, and thus to estimate the impact of fishing pressure on the marine ecosystem (Poos and Rijnsdorp, 2007; Mills *et al.*, 2007), or to understand fishermen's reactions to management measures (Vermard *et al.*, 2008) and to improve assessment of the impact of management plans (Lehuta *et al.*, 2013). Assessing the spatio-temporal distribution of vessels targeting fish populations also improves the understanding of the dynamics of fish resources (Bertrand *et al.* (2004); Poos and Rijnsdorp (2007)).

Modeling the dynamics of fishing vessels is classically approached by statistical analyses of landing declarations. This has a low spatial resolution (ICES statistical rectangle) (Hutton *et al.* (2004); Pelletier and Ferraris (2000)). Recently, the mandatory vessel monitoring system (VMS), for legal controls and safety (Kourti *et al.*, 2005), has led to massive acquisition of fishing vessel movement data, which offers new means of studying the spatio-temporal dynamics of fishermen. Data consist in geographical positions recorded at a more or less regular time step (less than 2 h for mandatory VMS data). The French Institute for the Exploitation of the Sea (IFREMER) has recently developed the RECOPESCA project with volunteer fishermen, whose vessel positions are recorded at a 15-min time step. Mechanistic mathematical models have long been used in ecological sciences to analyze movements and activity of different tracked animals (Bovet and Benhamou, 1988; Flemming *et al.*, 2006).

* *Correspondence to: Pierre Gloaguen, IFREMER, Ecologie et Modèles pour l'Halieutique, Rue de l'Ile d'Yeu, BP 21105, 44311 Nantes Cedex 03, France. E-mail: pierre.gloaguen@ifremer.fr*

a *IFREMER, Ecologie et Modèles pour l'Halieutique, BP 21105, 44311 Nantes Cedex 03, France*

b *AGROCAMPUS OUEST, UMR 985 Ecologie et Santé des Ecosystèmes, Ecologie Halieutique, CS84215, 35042 Rennes Cedex, France*

c *IFREMER, Sciences et Technologies Halieutiques, Technopôle Brest Iroise - BP 70, 29280 Plouzané*

d *IFREMER, Channel and North Sea Fisheries Departent, 150 Quai Gambetta, BP699, 62321 Boulogne sur mer, France*

e *AGROPARISTECH, UMR MIA 518, F-75231 Paris, France*

A key issue in behavioral ecology is the identification of the sequence of hidden (non-observed) behaviors from the analysis of the trajectory, such as foraging, searching, and migration. Similar questions are investigated in fisheries science, where the identification of different behaviors adopted by fishing vessels during a fishing trip (e.g. route toward fishing zone, fishing activity) is of interest to understand what drives fishing activities and fishing effort dynamics.

Hierarchical models are well suited for the analysis of trajectories and hidden sequence of behaviors from discrete position records. They first describe a non-observed time-behavioral process based on different behavioral states adopted by the individual and rules for switching from one to the other. The path is then modeled conditionally to the behavioral state. These models are commonly called state space models and (when the sequence of hidden state satisfies the Markov property) hidden Markov models (HMM) (Patterson *et al.*, 2009; Langrock *et al.*, 2012; Jonsen *et al.*, 2013). In animal ecology, these models were already used to describe the path of different animals such as elk (Morales *et al.*, 2004) or seals (Jonsen *et al.*, 2005). The application of those tools to infer fishing vessel activity has received comparatively little attention ((Vermard *et al.*, 2010), (Walker and Bez, 2010), or (Peel and Good, 2011)).

The model developed in this paper presents two substantial contributions to the models published thus far to analyze fishing vessel trajectories.

First, keeping the Markovian structure for the hidden behavior of the vessel, we propose to describe the vessel path via the modeling of the velocity process. The latter is modeled with a bivariate Gaussian first-order autoregressive process (AR); the parameters of which are conditioned by the behavioral state. Models published thus far have used scalar speed and turning angles to describe the vessel path conditionally upon the behavioral state. Those two variables are usually modeled without autocorrelation between time steps. For instance, the Bayesian model developed by Vermard *et al.* (2010) defines scalar speed and turning angles for each time step as drawn a priori in independent Gaussian and wrapped Cauchy distributions with parameters specific to the behavioral state. The AR bivariate velocity process allows a unique Gaussian structure to be used instead of two separated distributions for speed and turning angles. The AR process also includes the autocorrelation to capture the inertia in the velocity process (Gurarie *et al.*, 2009). This process has already been used in animal ecology to describe animal velocity, but, to the best of our knowledge, this work represents its first use in fisheries science to describe vessels path. It also offers computational facilities for the inferences.

Second, the model is fitted to original GPS record data from the RECOPESCA project (Leblond *et al.*, 2010). The RECOPESCA project is implemented by the French institute for the exploitation of the sea (IFREMER) to improve the assessment of the spatial distribution of catches and fishing. Although they concern a rather restricted number of fishing vessels, RECOPESCA data offer several advantages by comparison with mandatory VMS data. First, these data are recorded with a shorter time step than VMS data (a position every 15 min instead of every hour). Second, they are recorded with a highly regular time step (15 min $+/-$ 1 min). The finer time scale allows for a more accurate reconstruction of fishing vessel trajectories than do VMS data. Through an extensive simulation approach, Vermard *et al.* (2010) have already shown that a shorter acquisition time step would provide better inferences on the sequence of fishing vessel behavior. Bias induced by interpolating the trajectory with a straight line between two records would be lower than with an hour time step between two points (Skaar *et al.*, 2011). Furthermore, the regularity of these GPS records is essential for the formulation of the AR process hypothesis.

By contrast with the fisheries-related papers previously published that fostered a Bayesian approach ((Vermard *et al.*, 2010), (Walker and Bez, 2010)), a maximum likelihood estimation approach is adopted. The Baum–Welch (BW) algorithm, the expectation–maximization (EM) algorithm for HMM, is used to estimate parameters. The BW algorithm is then coupled with the Viterbi algorithm to estimate the hidden behavior sequence. In models previously published (Vermard *et al.* (2010); Walker and Bez (2010)), the sequence of hidden states has been inferred by using the posterior mode of marginal posterior distribution of each hidden state. By contrast, the Viterbi algorithm achieves estimation of the most credible sequence of behavioral states, which better accounts for the persistence in the sequence stemming from the Markovian (Rabiner, 1989).

The article is structured in the following way. In Section 2, we first detail the RECOPESCA data set. The methodological framework for the HMM coupled with an AR process is then presented. Estimation methods using EM and Viterbi algorithms are later detailed. Our simulation approach used to assess the performance of the estimation method and the sensitivity to alternative data configuration is presented at the end of Section 2. Results over simulations and RECOPESCA data are presented in Section 3. The ending section proposes a discussion on the adequacy of this modeling approach and some recommendations for future modeling of vessel dynamics.

2. MATERIAL AND METHODS

2.1. RECOPESCA data

Four trajectories associated with four different fishing vessels operating in the Channel with different fishing gear are considered to illustrate our modeling approach (Table 1). These four trajectories were extracted from the RECOPESCA database. For each trajectory, GPS positions in port and at sea were available. As the analysis only focuses on fishing vessel movement during fishing trips, we first removed positions in port based on logbooks (landing declarations). The positions were recorded at a regular time step (15 min $+/-$ 1). Selected trips last more than 12 h, ensuring enough observed positions for parameter identification. These four vessels belong to the demersal fishery for which the search for fish aggregations observed in pelagic fisheries (such as tuna fisheries, (Walker and Bez, 2010)) does not exist. Hence, only two behaviors are assumed along their path, ‘steaming’ for cruising and ‘fishing’ when they operate their gear.

Table 1. Technical details of the four studied trajectories

Trip	Duration	Vessel length	Gear
A	22h	12m	Dredges
B	14h	12m	Otter trawl
C	13h	13m	Trammel nets
D	107h	22m	Otter trawl

2.2. Describing the path with the velocity process

The observed vessel path, $X_0, \dots X_t$, is considered via a decomposition of the associated velocity process on its two dimensions V^p and V^r.

V^p is called the 'persistence' speed and corresponds to the tendency to persist in the previous direction. V^r is called the 'rotational' speed and corresponds to the tendency to turn. These two quantities are derived as follows:

$$V_t^p = V_t \cos(\psi_t) \tag{1}$$

$$V_t^r = V_t \sin(\psi_t) \tag{2}$$

where V_t is the average speed derived from positions X_{t-1} and X_t and ψ_t is the turning angle derived from X_{t-2}, X_{t-1}, and X_t, with $\psi_1 = 0$. Variables V^p and V^r model scalar speed and turning angles jointly instead of a different distribution for each of them (as in Vermard *et al.* (2010), or Walker and Bez (2010)). The bivariate velocity can be modeled using a unique Gaussian structure (Gurarie *et al.*, 2009), which is presented in the next section.

It is worth noting here that the velocity defined by Equations (1) and (2) is equivalent to the (linearly interpolated) observed trajectory (see Appendix for the explicit relation).

2.3. An autoregressive process ruled by a hidden Markov models

The vessel behavior is modeled by a hidden stochastic discrete time process. This is denoted by $S_0^t := S_0, ..., S_t$, where S_t is the state of the vessel at time t and takes values in the set of behavioral states noted $\mathcal{S} = \{1, 2\}$, with 1 for steaming and 2 for fishing.

This process is assumed to be a homogeneous Markov chain of the first order with a transition matrix $\Pi = (\Pi_{ik})_{i,k\in\mathcal{S}}$, that is,

$$\Pi_{ik} = \mathbb{P}(S_t = k | S_{t-1} = i) = \mathbb{P}\left(S_t = k | S_0^{t-2},\ S^{t-1} = i\right)$$

The initial distribution is assumed to be known and set at $\mathbb{P}(S_0 = 1) = 1$ (the vessel is steaming when leaving the harbor).

Conditionally to this hidden Markov chain, the velocity process is modeled by a mixture of two-dimensional AR processes (with respect to its decomposition in Equations (1) and (2)) and can be summarized as follows:

$$V_{t+1}^p | (S_{t+1} = i) = \eta_{p,i} + \mu_{p,i} V_t^p + \sigma_{p,i} \epsilon_{p,t} \tag{3}$$

$$V_{t+1}^r | (S_{t+1} = i) = \eta_{r,i} + \mu_{r,i} V_t^r + \sigma_{r,i} \epsilon_{r,t} \tag{4}$$

$$V_1^p = V_1,\ V_1^r = 0,\ \epsilon_{\cdot,t} \sim \mathcal{N}(0, 1)$$

where for each component (V^p or V^r) and state (1 or 2), the following are considered.

- η is a level parameter.
- μ is an autocorrelation parameter. Its existence is justified considering data from the four different trips (see Figure 1 for autocorrelation plots of trips A–D). It is important to note that it is well defined because of the time step regularity.
- σ^2 is a shape parameter and is the variance of the innovation process.

As in Gurarie *et al.* (2009), processes (3) and (4) are assumed to be independent. Even if this assumption seems unrealistic, data reveal a weak empirical correlation between those two variables.

AR processes as in (3) and (4) have a stationary distribution if $|\mu| < 1$ (Shumway and Stoffer, 2000). In this case, the expectation and the variance of the process V satisfy asymptotically:

$$\mathbb{E}(V) = \frac{\eta}{1 - \mu} \tag{5}$$

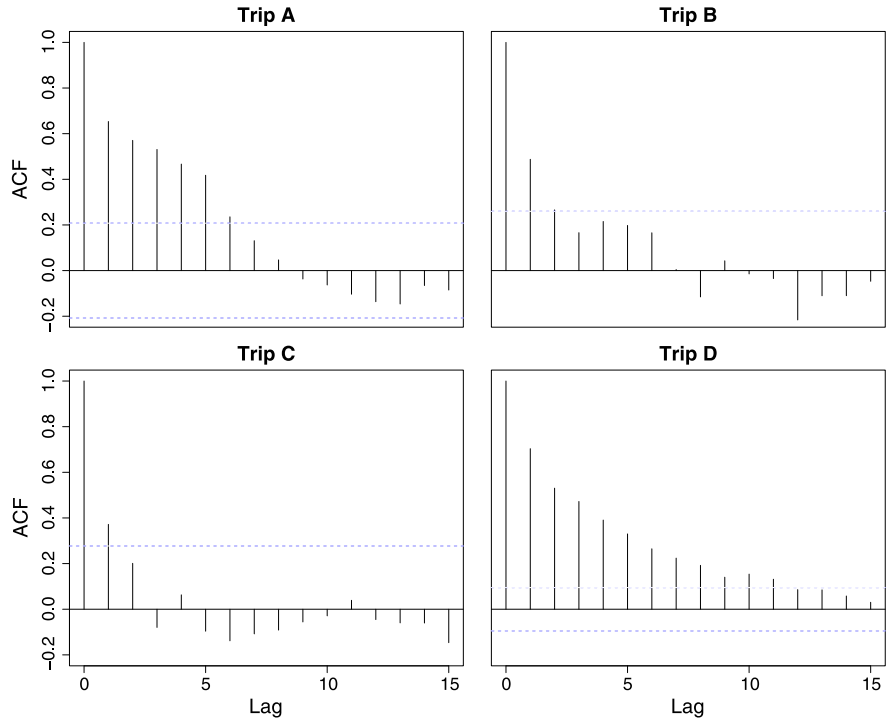

Figure 1. Autocorrelation function plots for trips A–D

Table 2. Parameter values for each simulation scenario. Matrix Π is identical for all scenarios. Asymptotic expectation and variance indicated are calculated from Equations (5) and (6) and rounded to the first digit. They are asymptotic and must be considered as indicators of how the parameters affect the different processes. n is the number of observations along the simulated trajectories

Scenario		η 1 2	μ 1 2	σ^2 1 2	$\mathbb{E}$ 1 2	$\mathbb{V}$ 1 2	n
1	p	$\begin{pmatrix} 6 & 1 \\ 0 & 0 \end{pmatrix}$			6 2		
	r				0 0		
2	p	$\begin{pmatrix} 6 & 2 \\ 0 & 0 \end{pmatrix}$	$\begin{pmatrix} 0 & 0.5 \\ 0 & 0.2 \end{pmatrix}$	$\begin{pmatrix} 1 & 0.5 \\ 0.5 & 0.1 \end{pmatrix}$	6 4	1 0.7	400
	r				0 0	0.5 0.1	
3	p	$\begin{pmatrix} 6 & 3 \\ 0 & 0 \end{pmatrix}$			6 6		
	r				0 0		
4	p		$\begin{pmatrix} 0 & 0.6 \\ 0 & 0.2 \end{pmatrix}$		6 2.5	1 0.8	
	r	$\begin{pmatrix} 6 & 1 \\ 0 & 0 \end{pmatrix}$		$\begin{pmatrix} 1 & 0.5 \\ 0.5 & 0.1 \end{pmatrix}$	0 0	0.5 0.1	400
5	p		$\begin{pmatrix} 0 & 0.8 \\ 0 & 0.2 \end{pmatrix}$		6 5	1 1.4	
	r				0 0	0.5 0.1	
6	p			$\begin{pmatrix} 2 & 0.5 \\ 1 & 0.1 \end{pmatrix}$		2 0.7	
	r	$\begin{pmatrix} 6 & 1 \\ 0 & 0 \end{pmatrix}$	$\begin{pmatrix} 0 & 0.5 \\ 0 & 0.2 \end{pmatrix}$		6 2	1 0.1	400
7	p			$\begin{pmatrix} 1 & 1 \\ 0.5 & 0.5 \end{pmatrix}$	0 0	1 1.3	
	r					0.5 0.5	
8	p						100
	r	$\begin{pmatrix} 6 & 1 \\ 0 & 0 \end{pmatrix}$	$\begin{pmatrix} 0 & 0.5 \\ 0 & 0.2 \end{pmatrix}$	$\begin{pmatrix} 1 & 0.5 \\ 0.5 & 0.1 \end{pmatrix}$	6 2	1 0.7	
9	p				0 0	0.5 0.1	50
	r						

$$\mathbb{V}(V) = \frac{\sigma^2}{1-\mu^2} \tag{6}$$

These asymptotic mean and variance are useful for interpreting characteristics of the velocity process. If the vessel stays long enough in a given behavior, the expectation for V^p and V^r could be derived from Equations 5 and (6).

2.4. Inference

The inference procedure consists in the estimation of both parameters and a reconstruction of the sequence of hidden states from observed positions. It requires two steps: (i) performing parameter estimation using the BW algorithm and (ii) estimating the most likely sequence of states using the Viterbi algorithm.

Considering two hidden states, the set of parameters to be estimated for this model is

$$\Theta = \{\Pi_{i\cdot},\ \eta_{p,i},\ \eta_{r,i},\ \mu_{p,i},\ \mu_{r,i}\ \sigma_{p,i},\ \sigma_{r,i}\ \}_{i\in\mathcal{S}}.$$

Therefore, 14 parameters are estimated (two for the transition matrix and $3 \times 2 \times 2$ for AR processes on the velocity). Computing the likelihood function is not possible within a reasonable time as it requires the integration over all possible hidden sequences ($2^{\#\text{Time steps}}$ possible paths). A classical approach is to find maximum likelihood estimates (MLE) $\hat{\Theta}$ via the BW algorithm, which is the EM algorithm derived for HMM ((Rabiner, 1989); (McLachlan and Krishnan, 1997)).

Considering the model described in the preceding text, both the expectation (E) step and the maximization (M) step can be computed analytically (details and proof are given in Appendix, and **R** codes (R Core Team, 2013) to perform the inference are available upon request).

The EM algorithm is an iterative algorithm to approach the MLE. The convergence criterion is reached when the log-likelihood increase is less than 0.01 between two successive iterations. A known problem of the EM algorithm is that, given a starting point, one can converge toward a local maximum of the likelihood. To ensure a global maximum is found, the algorithm was performed from 100 different starting points, keeping the result with the largest likelihood as $\hat{\Theta}$.

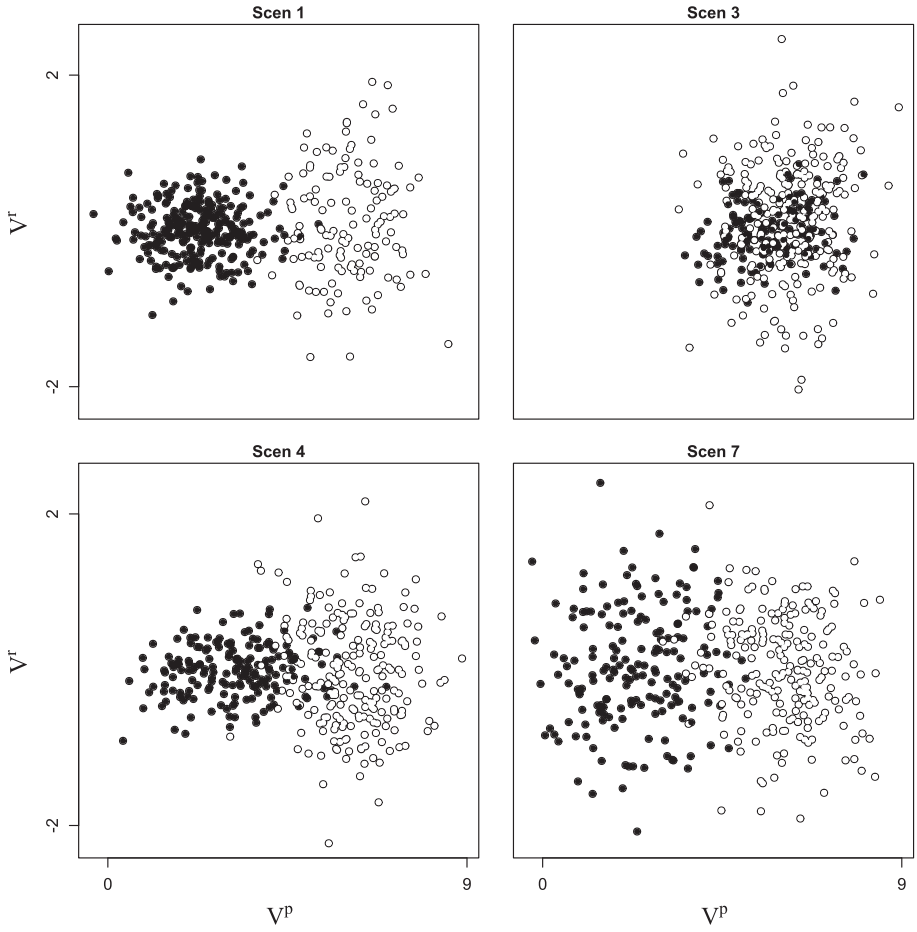

Figure 2. Simulated velocity processes for scenarios 1, 3, 4, and 7 (Table 2). The persistent speed V^p is represented along the x axis, and the turning speed is represented along the y axis. Black dots are for fishing, and white dots for steaming

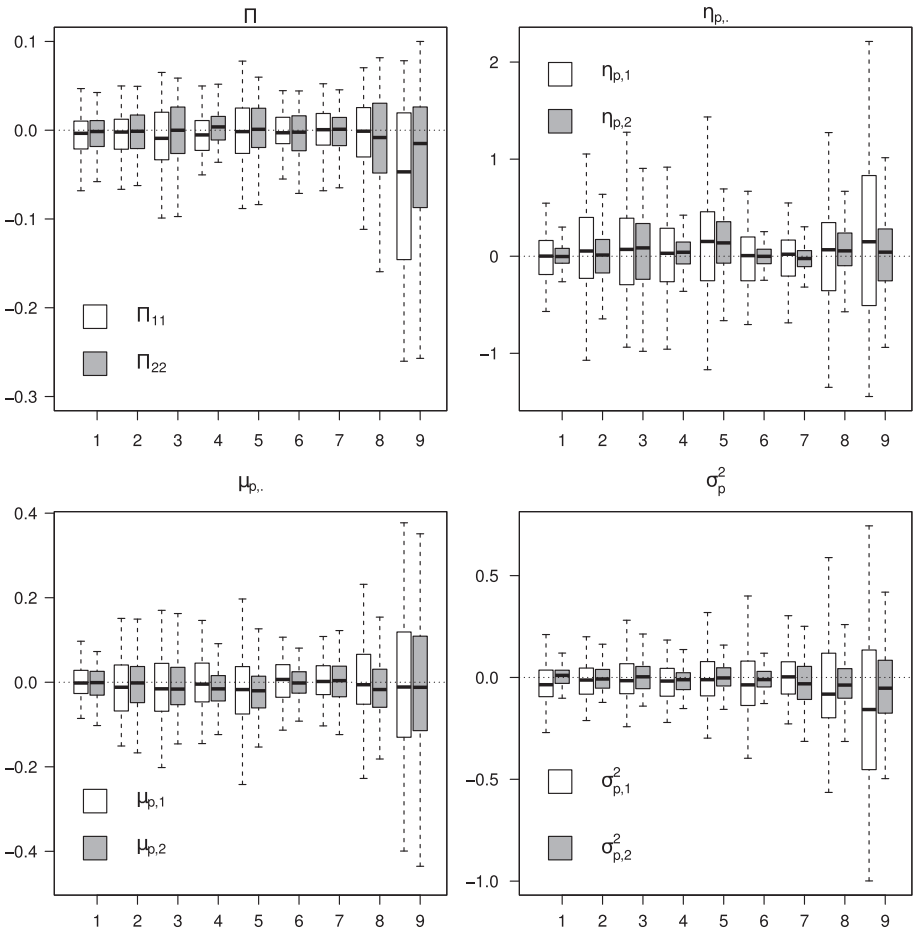

Figure 3. Estimation errors (estimated value minus the true value of each parameter, on the y axis) obtained for the nine simulation scenarios (x axis) presented in Table 2. Box plots represent the variability between the 100 simulations for each scenario. Only estimation errors for process V^p are presented. White and grey box plots are for parameter estimates in steaming and fishing, respectively. The whiskers represent here at most 1.5 times the interquartile range. Outliers are not plotted

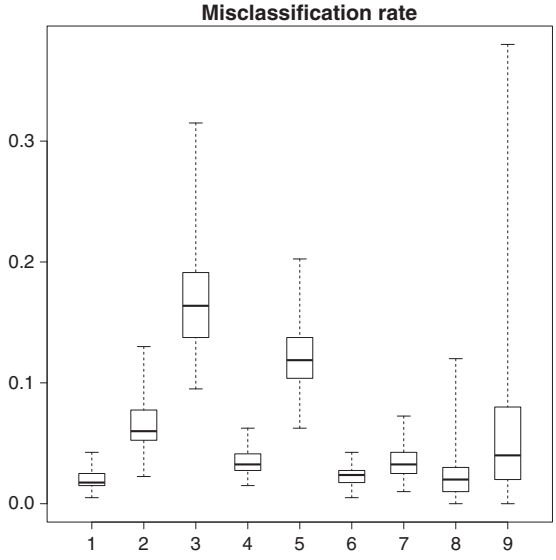

Figure 4. Box plots of misclassification rate from the Viterbi algorithm (on the y axis) for the nine simulation scenarios (on the x axis) presented in Table 2. Box plots represent the variability between the 100 simulations for each scenario. Whiskers height is at most two times the interquartile range. Outliers are not plotted

Once the MLE step is performed, the Viterbi algorithm is used to derive the most probable sequence of states, accounting for Markovian properties of the whole hidden sequence.

A parametric bootstrap procedure is used to assess the variance of $\hat{\Theta}$. The MLE is used to simulate M new trajectories as bootstrap samples on which MLE $(\hat{\Theta}_m)_{1\leq m\leq M}$ are re-estimated. Given these M re-estimations, empirical 95% confidence intervals are obtained for each parameter (obtaining central 95% values, (McLachlan and Krishnan, 1997; Efron and Tibshirani, 1993)).

To estimate the uncertainty over the state sequence estimation, the Viterbi algorithm is computed on the derived velocity process using each MLE $\hat{\theta}_m$. Formally, the Viterbi algorithm computes the most probable joint sequence of hidden states and observations:

$$\hat{S}^m = \text{argmax}_{s_0\ldots s_T}\left(p\left(s_0\ldots s_T,\ X_0\ldots X_T|\hat{\Theta}_m\right)\right) \tag{7}$$

The empirical probability of being in state 2 at time t is then computed as $\frac{\#\{m,\hat{S}_t^m=2\}}{M}$. 'Fishing' is attributed to the estimated state with the lowest mean for scalar speed, due to the fact that the vessel goes more slowly in that case.

The bootstrap is the most time-consuming part of the estimation.

2.5. Simulated scenarios

The performance of the estimation method is assessed through simulations of trajectories based on various scenarios mimicking different levels of contrast in the movement and activity characteristics. The nine scenarios considered are all variants of a baseline scenario (S1, described hereafter). However, they all share common properties on parameter values, chosen to fit the characteristics of the observed trajectories.

1. Movement does not privilege any turning direction. Then, we set $\eta_{r,1} = \eta_{r,2} = 0$, leading to an expected mean of V^r equal to 0 in both states.
2. When cruising, vessel goes faster than when fishing. Then we set $\frac{\eta_{p,1}}{1-\mu_{p,1}} > \frac{\eta_{p,2}}{1-\mu_{p,2}}$: the expected mean of persistent speed is greater when steaming than when fishing.
3. At a 15-min time step, the time spent in each step is expected to be large. Then, diagonal terms Π_{11} and Π_{22} of the transition matrix Π are large relative to the antidiagonal terms, meaning that the probability of staying in the same behavioral state is large relative to the probability of shifting. This matrix is common to all scenarios and set at $\Pi = \begin{pmatrix} 0.9 & 0.1 \\ 0.1 & 0.9 \end{pmatrix}$.

Trips are simulated following nine scenarios, with various degrees of mixture between states and with various numbers of observations (see Table 2 for detailed values).

- *Scenario 1:* (baseline scenario). The contrast between $\eta_{p,1}$ $(= 6)$ and $\eta_{p,2}$ $(= 1)$ is large, as is the difference in autocorrelation parameters ('Steaming' state is uncorrelated while 'Fishing' state is positively correlated).

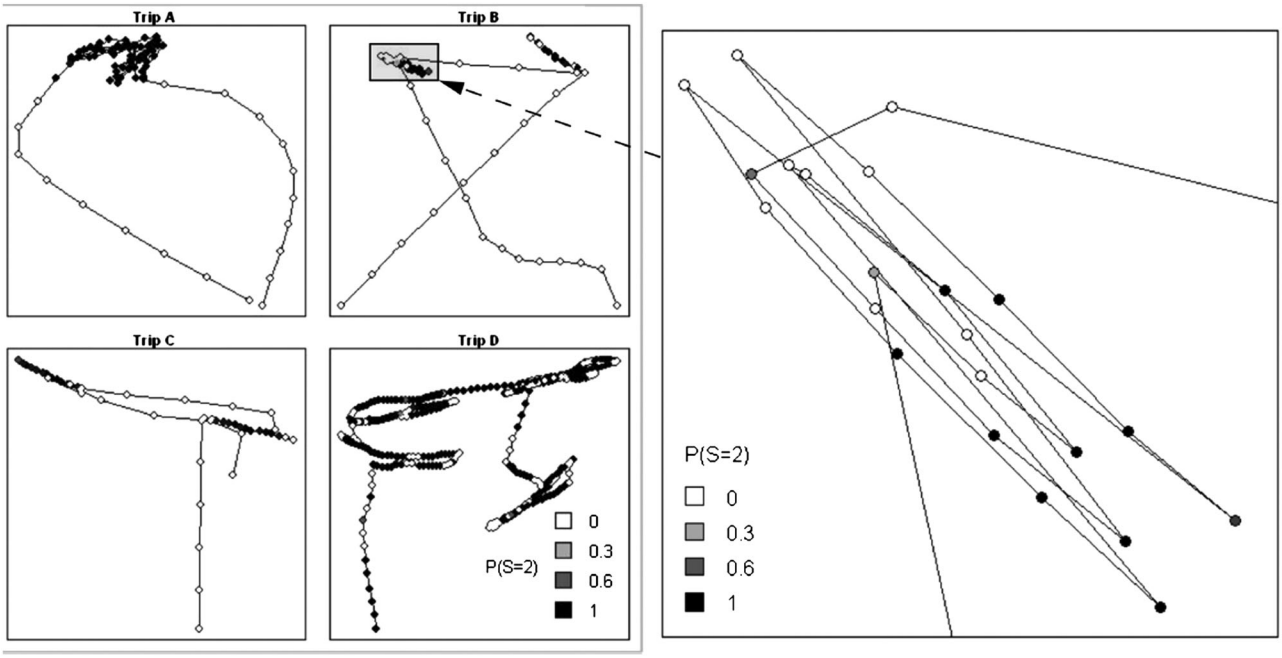

Figure 5. The four trajectories observed from RECOPESCA position records. Estimated probabilities of being in state 2 (fishing) for each trip (A, B, C, and D) are plotted, from 0 (white dots) to 1 (black dots). The right panel is a zoom over a specific zone of trip B (shaded rectangle). Estimations for trip D cannot be interpreted as steaming/fishing (see text)

- *Scenarios 2–3:* $\eta_{p,2}$ increases from 1 (scenario 1) to 2 (scenario 2) and 3 (scenario 3), resulting in an increase in the asymptotic expectation of V^p in state 2. Therefore the contrast in the expected asymptotic speed between state 1 and 2 decreases.
- *Scenarios 4–5:* $\mu_{p,2}$ increases from 0.5 (scenario 1) to 0.6 (scenario 4) and 0.8 (scenario 5), resulting in an increase in the asymptotic expectation of V^p in state 2. Therefore, the contrast in the expected asymptotic speed between states 1 and 2 decreases. Moreover, the asymptotic variance of process V^p increases in state 2.
- *Scenarios 6–7:* In scenario 6, $\sigma^2_{p,1}$ and $\sigma^2_{r,1}$ increase from 1 and 0.5 (scenario 1) to 2 and 1, respectively, resulting in a higher asymptotic variance in state 1. In scenario 7, $\sigma^2_{p,2}$ and $\sigma^2_{r,2}$ increase from 0.5 and 0.1 (scenario 1) to 1 and 0.5, respectively, resulting in a higher asymptotic variance in state 2.
- *Scenarios 8–9:* The number of observations is shortened from 400 points (scenario 1) to 100 points in scenario 8 and 50 points in scenario 9. Four hundred points and 50 points would represent, respectively, 100 and 12 h data considered and were the maximal and the minimal lengths of trajectories considered.

3. RESULTS

3.1. Results on simulated scenarios

Nine scenarios (corresponding to nine sets of parameters), mimicking different levels of contrast in the movement and activity characteristics, were simulated.

For each scenario, 100 trajectories are simulated, thus providing 100 parameter estimates and 100 estimations of the most credible sequence of behavioral states derived from the Viterbi algorithm.

Examples of velocity process obtained with parameters of scenarios 1, 3, 4, and 7 are represented on Figure 2. These scatter plots highlight the different degrees of mixture between the two states, depending on the scenario.

Knowing the true value of each parameter, estimation errors are computed and summarized using box plots (Figure 3). Results are shown only for process V^p, as trends are similar on process V^r (results for V^r are shown in Appendix). Moreover, as the true sequence of behavioral states is known, a misclassification rate is also computed and displayed using box plots (Figure 4).

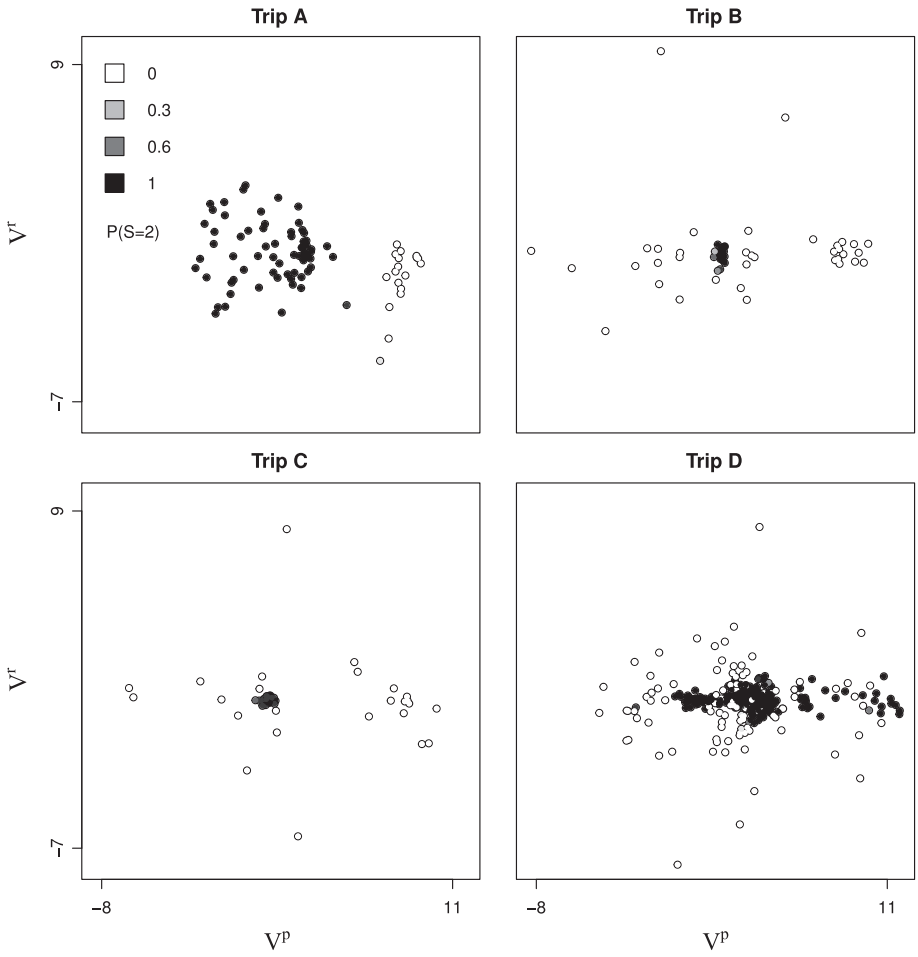

Figure 6. Velocity process for each trip (A, B, C, and D). The persistent speed V^p is represented along the x axis, and the turning speed is represented along the y axis. The estimated probability of being in state 2 (fishing) for each trip is plotted, from 0 (white dots) to 1 (black dots)

- *Scenarios 1–3:* For all parameters, the width of the box plots increases from scenarios 1 to 3, revealing that decreasing the contrast between $\eta_{p,1}$ and $\eta_{p,2}$ has a negative impact on the estimation of all parameters (Figure 3). Moreover, the misclassification rate of the behavioral states is also increased. Even if it remains low for scenarios 1 and 2, it increases for scenario 3 (more than 50% of the 100 simulations result in a misclassification rate greater than 0.15, Figure 4). Looking at Figure 2, this large misclassification rate is explained by the large degree of mixture between states in scenario 3.
- *Scenarios 1 and 4–5:* Increasing $\mu_{p,2}$ (while keeping $\eta_{p,1}$ and $\eta_{p,2}$ unchanged) increases estimation uncertainty over level and autocorrelation parameters η and μ (Figure 3). The misclassification rate also increases, with a low increase for scenario 4, and a larger one for scenario 5 (Figure 4). Indeed, there is an increase in the degree of mixture between states from scenario 1 to scenarios 4 and 5 (see Figure 2 for scenario 4).
- *Scenarios 1 and 6–7:* In scenario 6, increasing variance parameters $\sigma^2_{p,1}$ in state 1 slightly increases the uncertainty over the estimates of level and variance parameters $\eta_{p,1}$ and $\sigma^2_{p,1}$. The same effect is noticed in scenario 7 when variance parameters in state 2 increase (Figure 3). The misclassification rate remains stable between scenarios 1 and 6, but increases for scenario 7 as the processes in both states have in this case the same variance parameters (Figure 4). Indeed, there is an increasing in the degree of mixture between states from scenario 1 to scenario 7 (Figure 2).
- *Scenarios 1 and 8–9:* When the number of observations is shortened, estimation uncertainty increases for all parameters (Figure 3). Moreover, the misclassification rate is also impacted, getting worse as the observation length gets shorter (Figure 4). Looking at estimates of $\hat{\Pi}_{22}$ (for instance, the same can happen for $\hat{\Pi}_{11}$), it is worth noting that considering only 50 data points as in scenario 9 can lead to estimating $\hat{\Pi}_{22}$ close to 0. This results in the identification of only one behavioral state, and then in a large misclassification rate.

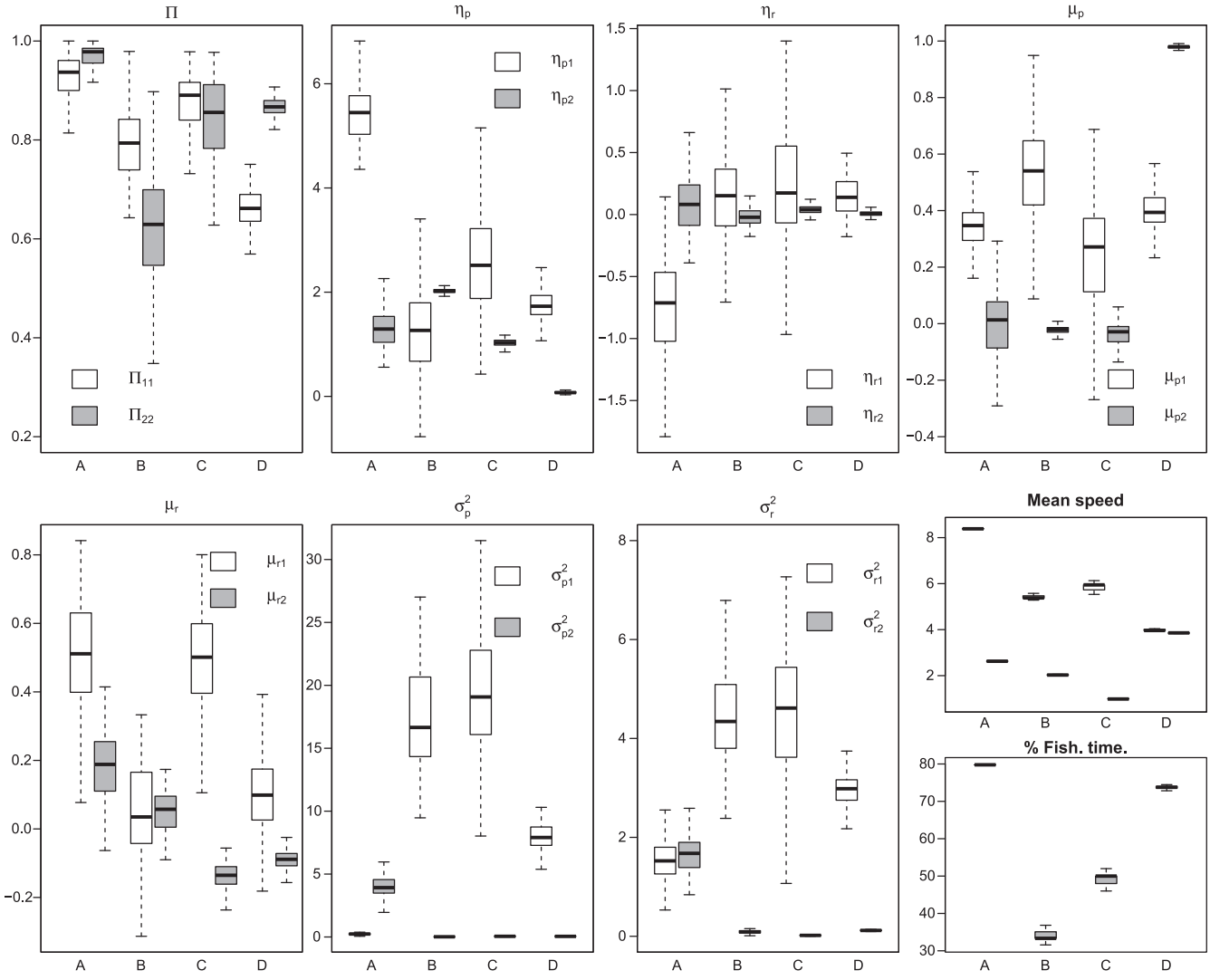

Figure 7. Estimations for parameters in each trip. The mean speed in each state and the proportion of time spent fishing are also presented. Box plots represent the variability over estimates as assessed by the bootstrap procedure. White boxes stand for state 1, and grey ones for state 2

More generally, it is worth noting that for all scenarios, estimations are unbiased. Moreover, except for scenario 7 where variance parameters are equal in both states, the variance of estimators is greater for state 1 parameters than for state 2 parameters, as the variance parameter is larger in the first state ($\sigma^2_{p,1} > \sigma^2_{p,2}$).

3.2. Results on RECOPESCA data

The four observed trajectories are represented on Figure 5 together with the estimated probabilities of fishing at each observed position of the vessel. Velocities associated with these trajectories are represented using scatter plots on Figure 6. The uncertainty in distinguishing the behavior of fishing from steaming is low for all trips, as the estimated probability of fishing is most of the time 0 or 1 (Figure 6). However, misidentification between fishing and steaming might occur at certain turning points (Figure 5).

Estimations for the 14 parameters, the estimated proportion of time spent fishing and the mean scalar speed in each behavioral state are presented on Figure 7. As expected, parameters Π_{11} and Π_{22} are high on those four trips showing a persistence to stay in a given activity. Parameters η_r are estimated quite close to 0, except on trip A. In this case, $\eta_{r,1}$ is slightly negative, stemming from the tendency of this vessel to always turn in the same direction during this trip.

The four trips have different patterns: trip A has an erratic fishing activity at low speed, trips B and C have a similar constant fishing activity pattern, with a constant speed and a steady course, and trip D is a mix between strongly autocorrelated speed patterns and abrupt changes.

For trip A, estimated steaming and fishing states are clearly separated on a scatter plot of velocity (Figure 6), steaming being concentrated at high values for V^p (large value of $\eta_{p,1}$) and fishing being more dispersed at lower values of V^p (smaller value of $\eta_{p,2}$). Steaming represents 20% of the vessel activity during the trip and takes place at high speeds (mean around 8.4 knots), while fishing represents 80% of the vessel activity, occurs at a lower speed (mean of 2.6 knots), and is more erratic.

For trips B and C, both components of the velocity process, V^p and V^r, have a very low variability (Figure 6). In those two cases, parameters $\mu_{.,2}$ and $\sigma^2_{.,2}$ are small, resulting in a process (while fishing) with a small variance. Fishing then occurs at constant speed and results in a steady course. By contrast, the steaming state is estimated with large variance parameters ($\sigma^2_{.,1}$ have large values). Then, as expected after simulation analysis, parameter estimates of V^p and V^r for this behavioral state are more uncertain.

For Trip D, estimated steaming and fishing states are more mixed. Parameter estimates show a low uncertainty for fishing and a larger uncertainty for steaming. Trip D has more than 400 observed positions; therefore, given the results of the simulation analysis, the uncertainty of state 1 estimated parameter can be associated to with large variance in the process (large values for $\sigma^2_{.,1}$). Moreover, state 2 is here characterized by a parameter $\mu_{p,2}$ very close to 1, translating a highly autocorrelated V^p process in state 2. Interpretation of results obtained in trip D is questionable. Behaviors are mixed along the whole trajectory and seem unrealistic in terms of steaming/fishing (Figure 5). Figure 8 presents the scalar speed process for this particular trip. It shows that scalar speed does not differentiate the two states: both fishing and steaming present high speed values. However, this is not realistic considering the vessel can operate an otter trawl at 10 knots. Actually, the two states are likely separated out based on the magnitude of autocorrelation in the V^p process, state 2 being highly autocorrelated ($\mu_{p,2} \approx 1$) with a small variance (low $\sigma^2_{p,2}$) and state 1 being less autocorrelated and with a larger variance parameter (larger $\sigma_{p,1}$). On Figure 8, the autocorrelated process corresponds to a portion of sine waves, and the other state is noise.

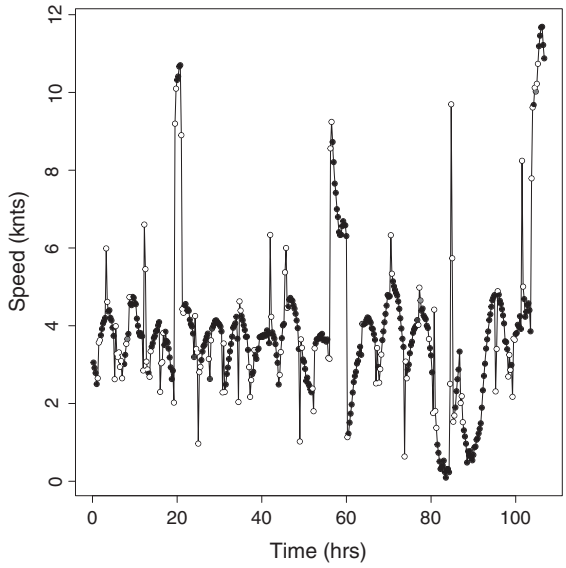

Figure 8. Scalar speed process for trip D, with estimated states. Probability of being in state 2 is plotted from 0 (white dots) to 1 (black dots). The model seems to fail to attribute behavioral state with respect to the speed, as the scalar speed is high for both states. The state attribution here is based on autocorrelation (black dots and white dots for high and low autocorrelation, respectively)

4. DISCUSSIONS AND PERSPECTIVES

This paper provides a first application of an AR process coupled with a hidden Markov chain to describe the movement of fishing vessels. The AR process allows a general framework with Gaussian properties and interpretable parameters. In this paper, it is shown how the velocity process viewed as an AR process can be used in order to analyze fishing vessel trajectories.

Application to the RECOPESCA data set provides insights into the understanding of fishing vessel activities at a fine spatio-temporal scale. Results over the four studied vessels highlight differences within trajectories (different types of vessels and fishing activities, (Biseau, 1998)) and within-trajectories (between steaming and fishing) in terms of time series characteristics (means, variance, and autocorrelation) that can also be translated in terms of physical patterns (fast, slow, erratic, and steady). These results show that the fishing activity performed influences the estimates of the velocity process parameters.

The model considered here has two states, steaming and fishing, which could be similar to a 'migrating'/'foraging' pattern adopted for animals (Jonsen *et al.*, 2007), whereas a three-state model can be used (Vermard *et al.*, 2010; Walker and Bez, 2010). This was made possible thanks to a pre-treatment of the data that consists in removing positions in port (which could be associated with a third state) but also because each studied fishing vessel operates with suitable gear that do not require research or stopping phase. If a two-state model is realistic here, it could be more relevant in other cases to adopt a three or more state model for trips during which several types of fishing gear can be operated or several métiers can be performed. A model with 'transition' states can also be adopted to deal with problems due to time step acquisition, and specifying different parameters for each fishery (Peel and Good, 2011). It is to be noted that increasing the number of states would not add any difficulty to the method presented here (Jonsen *et al.*, 2013). A more challenging alternative to these choices would be to consider a state space model where the number of states is a parameter to be inferred.

However, one has to question how to interpret the autocorrelation. In this study, trip D (trip of a 22 me bottom trawler) is disentangled into two states, one associated with highly autocorrelated persistent speed ($\mu_{p,2} \simeq 1$) and one associated with a less correlated persistent speed. It is to be wondered whether the estimated autocorrelation is of interest for the user's purpose (here, knowing when the vessel is fishing) or whether it is the expression of an external factor that does not reveal the hidden behavior. In the case of trip D presented here, the autocorrelated pattern might be the consequence of external factors that should be removed to accurately identify the steaming/fishing sequences. In that specific case, tide currents might be a cause of autocorrelated patterns. This covariate should then be included for future modeling of the vessel movement.

Along with this application on data, a simulation approach on nine contrasted scenarios is performed. The importance of the duration of observation is established as already noted by Vermard *et al.* (2010): the longer the trajectory, the better the estimation. Then, we recommend that results be interpreted with caution when dealing with small vessels as they might not have long enough trips to reliable estimation. In order to apply this model to VMS mandatory data, this has to be taken into account as a potential problem. This model might not be well suited for VMS data as the linearity hypothesis would be even stronger at a 60 min time step, and the irregularity makes the use of an AR process inappropriate. Moreover, the simulation approach pointed out problems to identifying two behaviors when the contrast between them is too small. In practice, it implies that two 'similar' fishing activities (for instance, dredging and trawling) might not be distinguished).

The AR process presented here, which allows the introduction of autocorrelation into the velocity process, was used by Jonsen *et al.* (2005) to model seal movement but was never applied to the movement of fishermen, which is often compared to top predator movement (Bertrand *et al.*, 2007). This approach allows us to model the bivariate velocity process (which fully describes the trajectory) with a unique Gaussian structure instead of two separate distributions as in previous models (Vermard *et al.*, 2010; Walker and Bez, 2010).

An interesting point of the model is its link with continuous time models. It is known that an AR process (with $\mu > 0$) is a discrete version of the continuous time Ornstein–Ulhenbeck process (OUP), as long as the time step is regular (Johnson *et al.*, 2008). Indeed, this model can be seen as an OUP sampled at discrete time coupled with an HMM, and there exists a bijection between the MLE of a positively correlated AR process and MLE of the regularly sampled OUP (see Appendix for proof). The OUP is known to be the solution of a specific stochastic differential equation (SDE). SDEs are a general and challenging tool to model spatial trajectories (Brillinger, 2010), as its continuous time property allows us to deal with irregularity in data but also to integrate spatially continuous covariates that rule the individual dynamics. However, it is to be mentioned that SDEs would require having instantaneous velocities, and not averaged velocities as in the presented model. This is a line of research we wish to explore further to circumvent modeling difficulties of VMS mandatory data.

The parameter estimation is performed using the BW algorithm and the reconstruction of the hidden state sequence is achieved thanks to the Viterbi algorithm. In order to compare results, inferences were also performed in a Bayesian framework (not shown here) using Markov chain Monte Carlo methods, and estimates showed similar results. Actually, there are several techniques for estimating parameters in HMM (Jonsen *et al.*, 2013). Considering the AR process presented here, the BW algorithm does not need numerical techniques as the equations associated have analytic solutions. This allows the algorithm entirely to be coded without any toolbox, which is, in our opinion, an advantage, even if this might not be faster (Zucchini and MacDonald, 2009). The uncertainty about estimates is assessed using bootstrap methods instead of Fisher matrix, which could be given as an output of the algorithm. Bootstrap methods might take longer to compute (this is the most time-consuming part of the algorithm), but it does not rely on asymptotic assumptions, which would not stand for short trips presented here(Efron and Tibshirani, 1993; Zucchini and MacDonald, 2009).

Acknowledgements

This work was funded by the EU project VECTORS and La Région Pays de Loire. The authors thank national research network Pathtis funded by INRA for helpful discussion and two anonymous reviewers for comments, which greatly improved this paper. This work would not have been carried out without all contributors to the RECOPESCA project. The authors express their warm thanks to the leaders of this

project Patrick Berthou and Emilie Leblond and all the voluntary fishermen involved in this project. Finally, the authors thank Joan Sobota and Yannick Lanza for their rereading of the English.

REFERENCES

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

Bertrand S, Diaz E, Niquen M. 2004. Interactions between fish and fisher's spatial distribution and behaviour: an empirical study of the anchovy (*Engraulis ringens*) fishery of Peru. *ICES Journal of Marine Science* **61**(7):1127–1136.

Biseau A. 1998. Definition of a directed fishing effort in a mixed-species trawl fishery, and its impact on stock assessments. *Aquatic Living Resources* **11**(3):119–136.

Bovet P, Benhamou S. 1988. Spatial-analysis of animals movements using a correlated random-walk model. *Journal of Theoretical Biology* **131**(4):419–433.

Brillinger D. 2010. *Handbook of Spatial Statistics*, Chapman and Hall/CRC Handbooks of Modern Statistical Methods. CRC Press: Boca Raton, FL.

Efron B, Tibshirani R. 1993. *An Introduction to the Bootstrap*, Monographs on Statistics & Applied Probability. Chapman & Hall.

Flemming JEM, Field CA, James MC, Jonsen ID, Myers RA. 2006. How well can animals navigate? Estimating the circle of confusion from tracking data. *Environmetrics* **17**(4):351–362.

Gurarie E, Andrews RD, Laidre KL. 2009. A novel method for identifying behavioural changes in animal movement data. *Ecology Letters* **12**(5):395–408.

Hutton T, Mardle S, Pascoe S, Clark RA. 2004. Modelling fishing location choice within mixed fisheries: English North Sea beam trawlers in 2000 and 2001. *ICES Journal of Marine Science* **61**(8):1443–1452.

Johnson DS, London JM, Lea M-A, Durban JW. 2008. Continuous-time correlated random walk model for animal telemetry data. *Ecology* **89**(5):1208–1215.

Jonsen ID, Basson M, Bestley S, Bravington MV, Patterson TA, Pedersen MW, Thomson R, Thygesen UH, Wotherspoon SJ. 2013. State-space models for bio-loggers: A methodological road map. *Deep Sea Research Part II Topical Studies in Oceanography* **88–89**(SI):34–46.

Jonsen ID, Myers RA, James MC. 2007. Identifying leatherback turtle foraging behaviour from satellite telemetry using a switching state-space model. *Marine Ecology-Progress Series* **337**:255–264.

Jonsen ID, Flenming JM, Myers RA. 2005. Robust state-space modeling of animal movement data. *Ecology* **86**(11):2874–2880.

Kourti N, Shepherd I, Greidanus H, Alvarez M, Aresu E, Bauna T, Chesworth J, Lemoine G, Schwartz G. 2005. Integrating remote sensing in fisheries control. *Fisheries Management and Ecology* **12**(5):295–307.

Langrock R, King R, Matthiopoulos J, Thomas L, Fortin D, Morales JM. 2012. Flexible and practical modelling of animal telemetry data: Hidden Markov models and extensions. *Ecology* **93**(11):2336–2342.

Leblond E, Lazure P, Laurans M, Rioual C, Woerther P, Quemener L, Berthou P. 2010. The RECOPESCA project : A new example of participative approach to collect fisheries and in situ environmental data. *CORIOLIS Quarterly Newsletter* **37**:40–48.

Lehuta S, Petitgas P, Mahévas S, Huret M, Vermard Y, Uriarte A, Record NR. 2013. Selection and validation of a complex fishery model using an uncertainty hierarchy. *Fisheries Research* **143**(0):57 –66.

McLachlan GJ, Krishnan T. 1997. *The EM Algorithm and Extensions*, Vol. 274. Wiley: New York.

Mills CM, Townsend SE, Jennings S, Eastwood PD, Houghton CA. 2007. Estimating high resolution trawl fishing effort from satellite-based vessel monitoring system data. *ICES Journal of Marine Science* **64**(2):248–255.

Morales JM, Haydon DT, Frair J, Holsiner KE, Fryxell JM. 2004. Extracting more out of relocation data: building movement models as mixtures of random walks. *Ecology* **85**(9):2436–2445.

Patterson TA, Basson M, Bravington MV, Gunn JS. 2009. Classifying movement behaviour in relation to environmental conditions using hidden Markov models. *Journal of Animal Ecology* **78**(6):1113–1123.

Peel D, Good NM. 2011. A hidden Markov model approach for determining vessel activity from vessel monitoring system data. *Canadian Journal of Fisheries and Aquatic Sciences* **68**(7):1252–1264.

Pelletier D, Ferraris J. 2000. A multivariate approach for defining fishing tactics from commercial catch and effort data. *Canadian Journal of Fisgeries and Aquatic Sciences* **57**(1):51–65.

Poos J-J, Rijnsdorp AD. 2007. The dynamics of small-scale patchiness of plaice and sole as reflected in the catch rates of the dutch beam trawl fleet and its implications for the fleet dynamics. *Journal of Sea Research* **58**(1):100–112.

R Core Team. 2013. *R: A language and Environment for Statistical Computing*. R Foundation for Statistical Computing: Vienna, Austria.

Rabiner LR. 1989. A tutorial on hidden Markov models and selected applications in speech recognition. *Proceedings of the IEEE* **77**(2):257–286.

Skaar KL, Jorgensen T, Ulvestad BKH, Engas A. 2011. Accuracy of VMS data from Norwegian demersal stern trawlers for estimating trawled areas in the Barents sea. *ICES Journal of Marine Science* **68**(8):1615–1620.

Shumway RH, Stoffer DS. 2000. *Time Series analysis and Its Applications*, Springer Texts In Statistics. Springer-Verlag GmbH: New York, NY.

Walker E, Bez N. 2010. A pioneer validation of a state-space model of vessel trajectories (VMS) with observers' data. *Ecological Modelling* **221**(17):2008–2017.

Vermard Y, Marchal P, Mahevas S, Thebaud O. 2008. A dynamic model of the bay of biscay pelagic fleet simulating fishing trip choice: the response to the closure of the european anchovy (*engraulis encrasicolus*) fishery in 2005. *Canadian Journal of Fisheries and Aquatic Sciences* **65**(11):2444–2453.

Vermard Y, Rivot E, Mahevas 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**(15):1757–1769.

Zucchini Walter, MacDonald IL. 2009. *Hidden Markov Models for Time Series: An Introduction Using R*. CRC Press: Boca Raton, FL.

SUPPORTING INFORMATION

Additional supporting information may be found in the online version of this article at the publisher's web site.

