
An autoregressive model to describe fishing vessel movement and activity

Gloaguen Pierre ^{1,*}, Mahevas Stephanie ¹, Rivot Etienne ², Woillez Mathieu ^{2,3}, Guitton Jerome ², Vermard Youen ⁴, Etienne Marie-Pierre ⁵

¹ IFREMER, EMH, Rue de l'Île d'Yeu, BP 21105, 44311 Nantes Cedex 03, France

² AGROCAMPUS OUEST, Ecologie Halieutique, 65 rue de St-Brieuc, CS84215, 35042 Rennes Cedex, France

³ IFREMER, Sciences et Technologies Halieutiques, Plouzane, France

⁴ IFREMER, Channel and North Sea Fisheries Department, 150 Quai Gambetta, BP699, 62321 Boulogne sur mer, France

⁵ AGROPARISTECH, UMR MIA 518, F-75231 Paris, France

* Corresponding author : Pierre Gloaguen, email address : pierre.gloaguen@ifremer.fr

Abstract :

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 RECOPECA 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, RECOPECA, autoregressive process, Baum–Welch algorithm

1. INTRODUCTION

The understanding of the dynamics of fishing vessels is an essential tool to characterize spatial distribution of fishing effort on a fine spatial scale, 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 the assessment of the impact of management plans (Lehuta *et al.*, 2013). Another key issue addressed with models of fishing vessel dynamics concerns the understanding of the population dynamics derived from the spatio-temporal distribution of vessels targeting fish populations (Bertrand *et al.*, 2004; Poos and Rijnsdorp, 2007).

Modelling the dynamics of fishing vessels is classically approached by statistical analyses of landing declarations with 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 vessels' movement data which offer new means of studying VMS fishermen spatio-temporal dynamics. Data consist in geographical positions recorded at a more or less regular time step (less than two hours for mandatory VMS data) with low positioning errors. In addition to this, IFREMER developed the RECOPECA project with volunteer fishermen, whose vessels positions are recorded at a 15 minute time step. Mechanistic mathematical models have long been used in ecological sciences to analyse movements and behavior of different tracked animals (Bovet and Benhamou, 1988; Flemming *et al.*,

^aIFREMER, *Ecologie et Modèles pour l'Halieutique*, Rue de l'Île d'Yeu, BP 21105, 44311 Nantes Cedex 03, France

^bAGROCAMPUS OUEST, UMR 985 *Ecologie et Santé des Ecosystèmes*, *Ecologie Halieutique*, 65 rue de St-Brieuc, CS84215, 35042 Rennes Cedex, France

^cIFREMER, *Sciences et Technologies Halieutiques*, Technopôle Brest Iroise - BP 70, 29280 Plouzané

^dIFREMER, *Channel and North Sea Fisheries Department*, 150 Quai Gambetta, BP699, 62321 Boulogne sur mer, France

^eAGROPARISTECH, UMR MIA 518, F-75231 Paris, France

*. Correspondence to: E-mail pierre.gloaguen@ifremer.fr

2006). 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, research, migration. Similar questions are investigated in fisheries science, where the identification of different behaviors adopted by fishing vessels during a fishing trip (route towards fishing zone, fishing activity...) is of interest to understand what drives fishing activities and fishing effort dynamics. The associated mathematical models are hierarchically structured. 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 modelled conditionally to the behavioral state. These models are commonly called State Space Models (SSM) and (when the sequence of hidden state satisfies the Markov property) Hidden Markov Models (HMM) and have proved their usefulness for both ecology and fisheries science (Patterson *et al.*, 2009; Langrock *et al.*, 2012; Jonsen *et al.*, 2013). In animal ecology, these models are used to describe the path of different animals such as elks (Morales *et al.*, 2004) or seals (Jonsen *et al.*, 2005), for instance. Similar models are used in fisheries science to describe fishing vessels activity (Vermard *et al.*, 2010, Walker and Bez, 2010, or Peel and Good, 2011). So far, in fisheries science, the vessel's path is described using scalar speed and turning angles as the modelled variables. These two variables separately using Gaussian and Wrapped Cauchy distributions, for instance (Vermard *et al.*, 2010).

The primary goal of this paper is to investigate the possibility of describing fishing vessel's path with a HMM coupled to an autoregressive (AR) process. Keeping the Markovian structure for the hidden behavior of the vessel, we propose to describe the vessel's path, conditionally to this behavior, via the modelling of vessel's velocity, using an AR process. The bivariate velocity process allows the use of a unique Gaussian structure (Gurarie *et al.*, 2009) instead of two separated distributions for speed and turning angles. This variable has already been used in animal ecology to describe animal's path, but, to the best of our

knowledge, this work represents its first use in fisheries science to describe vessel's path. To estimate parameters of the model, a maximum likelihood estimation approach is adopted. The Baum Welch algorithm is used to estimate parameters and then coupled to the Viterbi algorithm to estimate the hidden behavior sequence.

We illustrate the strength and limitations of the approach by fitting the model to GPS records issued from the RECOPECA project (Leblond *et al.*, 2010), implemented by IFREMER to improve the assessment of the spatial distribution of catches and fishing. Although they concern a rather restricted number of fishing vessels, RECOPECA 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 minutes instead of 1 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 VMS data. In particular, 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 recording is essential to formulate the AR process hypothesis.

The article is structured in the following way. In section 2, we detail the RECOPECA data set, followed by the theoretical and methodological framework including the model's description, the inference algorithm and the simulation approach to assess the performance of the method. Results over simulations and real data are presented in section 3. The ending section proposes a discussion on the adequacy of this modelling approach and some recommendations for future modelling of vessels dynamics.

2. MATERIAL AND METHODS

2.1. RECOPECA data

Four trajectories associated with four different fishing vessels operating in the Channel with different fishing gears are considered to illustrate our modelling approach (see Table 1). These four trajectories were extracted from the RECOPECA data base. For each trajectory, GPS positions in port and at sea were available. As the analysis only focus on fishing vessel movement during fishing trips, we first removed positions in port based on logbooks (landings declarations). The positions were recorded at a regular time step (plus or minus 1 minute). Selected trips last more than 12 hours, ensuring enough observed positions for parameters identification. These four vessels belong to the demersal fishery for which the research of fish aggregations observed in pelagic fisheries (such as thuna 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.

Despite their differences, common characteristics are assumed for these trajectories : 1) movement is mainly in a straight line with no turning direction privileged ; 2) when cruising, vessel goes faster than when fishing ; 3) at a 15 minutes time step, there is mainly persistence in a state (states occur in sequence). While animals usually show a more erratic movement when foraging than when cruising (Morales *et al.*, 2004), for fishing vessels, the fishing phase can be either erratic or not depending on the operated gear. For instance, it is sometimes erratic for dredging whereas it is more linear for trammelling or otter trawling. Therefore, this erratic behavior might not always be appropriate to distinguish between fishing and cruising states.

[TABLE 1 about here.]

2.2. Describing the path with the velocity process

The observed vessel's 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 maintain 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 1 and 2 model jointly scalar speed and turning angles instead one different distribution for each of them (like in and Walker and Bez, 2010). The bivariate velocity can be modelled using a unique Gaussian structure (Gurarie *et al.*, 2009), that is presented on 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 the appendix for the explicit relation).

2.3. An AR process ruled by a HMM

The vessel's behavior is modelled by a hidden stochastic discrete time process noted $S_0^t := S_0, \dots, S_t$, where S_t is the state of the vessel at time t , and takes values in the set of behavioral states noted $\mathcal{S} = \{1, \dots, I\}$.

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

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

In our specific study, only two states are considered, $\mathcal{S} = \{1, 2\}$, 1 standing for steaming, 2 for fishing.

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

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

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

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

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

where, for each component (V^p or V^r) and state (1 or 2) :

- 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's important to note that it is well defined because of the time step regularity.
- σ^2 is a shape parameter, it is the noise 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.

It is known that an AR process as in (3) and (4) has a stationary distribution if $|\mu| < 1$ (Shumway and Stoffer, 2000). In this case the expectation and the variance of the process V

satisfy asymptotically :

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

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

These asymptotic equalities can be useful in order to interpret 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). These asymptotic equalities can then be used as first approximations for the expectation and the variance of the process in a given state.

[FIGURE 1 about here.]

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 1) performing parameter estimation using the Baum Welch algorithm, 2) estimating the most likely sequence of states using the Viterbi algorithm.

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

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

When $J = 2$, 14 parameters are estimated (2 for the transition matrix, and $3 \times 2 \times 2$ for AR processes parameters). Computing the likelihood function in this case is not possible within a reasonable time as the computation of the likelihood requires the integration over all possible hidden sequences ($2^{\#\text{Observations}}$ possible paths). A classical approach is to find Maximum Likelihood Estimators (MLE) $\hat{\Theta}$ via the Baum Welch algorithm, which is the

Expectation Maximization (EM) algorithm derived for Hidden Markov Models (Rabiner, 1989; McLachlan and Krishnan, 1997).

Considering the model described above, both the Expectation (E) step and the Maximization (M) step can be computed analytically (details and proof are given in the appendix, and **R** codes (R Core Team, 2013) to perform the inference are available on demand).

The convergence criterion is reached when the log likelihood increase is less than 0.01. A known problem of the EM algorithm is that, given a starting point, one can converge towards 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 $\hat{\Theta}$.

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 (getting 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

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

The empirical probability of being in state 2 at time t is then computed as $\frac{\#\{m, \hat{S}_t^m=2\}}{M}$.

Working on real data, state 2 (standing for "fishing") is attributed to the estimated state with the lowest mean for scalar speed, due to the fact that the vessel goes slower in that case.

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

2.5. Simulations

The performance of the estimation method is assessed through simulations of trajectories based on various scenarios mimicking different levels of contrast in the movement characteristics of the two behavioral states. For the model with two different states (14 parameters), parameters used for simulation are restricted to values consistent with characteristics of the observed trajectories. For instance, we noticed that

- 1) $\eta_{r,1} = \eta_{r,2} = 0$: the movement of vessels is mainly in a straight line, both while steaming and fishing, and does not privilege any turning direction, which implies the level parameter to be null ;
- 2) $\frac{\eta_{p,1}}{1-\mu_{p,1}} > \frac{\eta_{p,2}}{1-\mu_{p,2}}$: the asymptotic component of persistent speed should be greater when steaming than when fishing ;
- 3) diagonal terms Π_{11} and Π_{22} of the transition matrix Π are large. This matrix is common to all scenarios and set to $\Pi = \begin{pmatrix} 0.9 & 0.1 \\ 0.1 & 0.9 \end{pmatrix}$.

Trips are simulated following nine scenarios with various degrees of mixture between states and with different lengths (number of time steps) (see Table 2 for detailed values).

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

Scenario 2-3 $\eta_{p,2}$ increases from 1 (scenario 1) to 2 (scenario 2) and 3 (scenario 3), resulting in an increase of the asymptotic expectation of V^p in state 2. Therefore the contrast in the expected asymptotic speed between state 1 and 2 decreases.

Scenario 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 of the asymptotic expectation of V^p in state 2. Therefore the contrast in the expected asymptotic speed between state 1 and 2 decreases. Moreover, the asymptotic variance of process V^p increase in state 2.

Scenario 6-7 In scenario 6, $\sigma_{p,1}^2$ and $\sigma_{r,1}^2$ increase from 1 and 0.5 (scenario 1) to 2 and 1 respectively, resulting to a higher asymptotic variance in state 1. In scenario 7, $\sigma_{p,2}^2$ and $\sigma_{r,2}^2$ increase from 0.5 and 0.1 (scenario 1) to 1 and 0.5 respectively, resulting to a higher asymptotic variance in state 2.

Scenario 8-9 The length of the observation is shortened from 400 points (scenario 1) to 100 points in scenario 8 and 50 points in scenario 9. Lengths of 400 and 50 points would represent respectively 100 and 12 hours data considered and were the maximal and the minimal length of trajectories considered.

[TABLE 2 about here.]

3. RESULTS

3.1. Results on simulations

For each set of parameters, 100 trajectories are simulated, thus providing 100 parameter estimates. Examples of velocity process obtained with parameters of scenarii 1, 3, 4 and 7 are represented using scatter plots 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 . Moreover, as the true sequence of behavioural states is known, a misclassification rate is also computed and displayed using box plots (Figure 4).

Box plots results highlight performances of the parameters estimation method and the Viterbi algorithm, which are now detailed for the different simulation scenarios :

Scenario 1-3 For all parameters, the width of the box plots increases from scenarios 1 to 3, revealing that an increase proximity 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 states estimation is also increased. Even if it remains low for scenarios 1 and 2, it increases for scenario 3 (more than 50% of states estimation have a misclassification rate greater than 0.15, Figure 4). Looking at Figure 2, this large misclassification rate can be explained by the large degree of mixture between states in scenario 3.

Scenario 1, 4-5 When $\eta_{p,1}$ and $\eta_{p,2}$ are not changed, the increase of $\mu_{p,2}$ increases estimation's 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).

Scenario 1, 6-7 In scenario 6, increasing noise parameters $\sigma_{p,1}^2$ in state 1 increases lightly the uncertainty for the estimations of level and noise parameters $\eta_{p,1}$ and $\sigma_{p,1}^2$. The same trend can be noticed in scenario 7 when noise parameters in state 2 increase (Figure 3). The misclassification rate remains stable between scenario 1 and 6, but increases for scenario 7 as the processes in both states have in this case same noise parameters (Figure 4). Indeed, there is an increasing in the degree of mixture between states from scenario 1 to scenario 7 (Figure 2).

Scenario 1,8-9 When the length is shortened, estimation's uncertainty increases for all parameters, the increase becomes larger from scenario 8 (100 points) to scenario 9 (50 points) (Figure 3). Moreover, the misclassification rate is also impacted, getting worse as the observation's 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 in scenario 9, this parameter is sometimes estimated close to 0. This results in the identification of only one behavioral state, and then a large misclassification rate.

More generally, it is worth noting that for all scenarios, estimations are unbiased. Moreover, except for scenario 7 where noise parameters are equal in both states, the variance of estimators is greater for state 1 parameters than for state 2 parameters, as the noise

parameter is larger in the first state ($\sigma_{p,1}^2 > \sigma_{p,2}^2$).

[FIGURE 2 about here.]

[FIGURE 3 about here.]

[FIGURE 4 about here.]

3.2. Results on data

The four observed trajectories with estimated probabilities of fishing at each observed position of the vessel (with a focus on a particular section of trip B) are represented on Figure 5. Velocities associated to these trajectories are represented using scatter plots on Figure 6. One can see that 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 (see Figure 6). However, misidentification between fishing and steaming might occur at certain turning points (see Figure 5).

Estimations for the 14 parameters, estimated proportion of time spent fishing and 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.

Results obtained for parameters estimates can be linked, for each trip, to the velocity scatter plot.

For trip A, estimated steaming and fishing states are clearly separated on 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's activity and takes place at high speeds (mean around 8.4

knots), while fishing represents 80% of the vessel's activity, it occurs at a lower speed (mean of 2.6 knots) and is rather erratic.

For trips B and C have a similar estimated fishing state pattern, which is really concentrated (Figure 6). Indeed, 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 is in a straight line. The steaming state are estimated with large noise parameters ($\sigma_{.,1}^2$ have large values), then, as expected after simulation analysis, parameters estimates for this behavioral state are more uncertain.

For Trip D, estimated steaming and fishing states are more mixed. Parameters estimates show a small 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 a large noise in the process (large values for $\sigma_{.,1}^2$). Moreover, state 2 is here characterized by a parameter $\mu_{p,2}$ really close to 1, traducing a highly autocorrelated V^p process in state 2. The fit of the model to 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 discriminate the two states : both fishing and steaming present high speed values. However it is not realistic considering the vessel can operate an otter trawl at 10 knots. Actually, the 2 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 noise (low $\sigma_{p,2}^2$, and state 1 being less autocorrelated and with a larger noise parameter (larger $\sigma_{p,1}$). On Figure 8, the autocorrelated process corresponds to a portion of sine waves, the other state is noise.

To conclude, the four trips have different patterns, trip A has an erratic fishing activity at low speed, trip 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

brutal changes.

[FIGURE 5 about here.]

[FIGURE 6 about here.]

[FIGURE 7 about here.]

[FIGURE 8 about here.]

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. This approach allows to model the bivariate velocity process (which fully describe the trajectory) with a unique Gaussian structure instead of two separated distributions as it was done before (Vermard *et al.*, 2010; Walker and Bez, 2010). 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 this point of view, in order to analyse fishing vessels trajectories. Indeed results over the four studied vessel's highlight differences inter-trajectories (different types of vessels and fishing activities Biseau, 1998) and intra-trajectories (between steaming and fishing) in terms of time series characteristics (means, variance, autocorrelation) that can also be translated in terms of physical patterns (fast, slow, erratic, steady). However, as for all time series study, one has to question how to interpret the estimated autocorrelation. In this study, the trip of a 22 meters bottom trawler is disentangled into two states, one associated to highly autocorrelated persistent speed ($\mu_{p,2} \simeq 1$) and one associated one associated to a less correlated persistent speed. It is to wonder 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 the trip D presented here, the autocorrelated pattern might be the consequence of external factors (surface currents for instance) that should be removed to accurately identify the steaming/fishing sequences.

The model considered here has two states, steaming and fishing, that could be similar to a "migrating"/"foraging" pattern adopted for animals (Jonsen *et al.*, 2007), whereas a three-states 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 but also because each studied fishing vessel operates with suitable gears that do not require research or stopping phase. If a two-states model is realistic here it could be more relevant in other cases to adopt a three or more states for trips during which several gears can be operated or several métiers can be practised . 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 note that increasing the number a state would not add any difficulty to the method presented here (Jonsen *et al.*, 2013). A challenging alternative to these choices would be to consider a state space model where the number of states is a parameter to infer.

The parameter estimation is performed using the Baum Welch algorithm and the reconstruction of the hidden state sequence is achieved thanks to the Viterbi algorithm. In order to compare results, the estimation was also made in a Bayesian framework (not shown here), with MCMC methods, using the Jags software, and estimates showed similar results. Actually, there are several techniques to estimate parameters in Hidden Markov Models (see Jonsen *et al.*, 2013). Considering the AR process presented here, the Baum Welch algorithm does not need numerical techniques as the equations associated have analytic solutions. This allows to code the algorithm entirely without any toolbox, which

is, in our opinion, an advantage, even if this might not be faster (Zucchini and MacDonald, 2009). The uncertainty of estimates is computed using bootstrap methods instead of Fisher matrix, that could be given as an output of the algorithm. Bootstrap methods might be longer to compute (it's the most time-consuming part of the algorithm) but it does not rely on asymptotic assumptions which would not stand here for at least 3 of the 4 trips presented (Efron and Tibshirani, 1993; Zucchini and MacDonald, 2009).

A simulation approach is performed to assess the performance of the model and the accuracy of the estimation for various realistic set of parameters values (called scenarios). The importance of the duration of observation is also established : the longer the trajectory, the better the estimation. Then, when dealing with small vessels as they might not have long trips enough to enable good estimation. Moreover, the simulation approach showed problems to identify two behaviors when the contrast between them is too small. In practice, it implies that two "similar" activities (for instance dredging and trawling) might not be distinguished.

An interesting point of the model presented here 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 a HMM. 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 to deal with irregularity in data but also to integrate spatially continuous covariates that rule the individual dynamics. This is a line of research we wish to plore further to circumvent modelling difficulties of VMS mandatory data.

ACKNOWLEDGMENTS

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 RECOPECA 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, authors thanks Yannick Lanza for his relecture of the english.

RÉFÉRENCES

- 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*, chapter 26. Chapman and Hall/CRC Handbooks of Modern Statistical Methods, CRC Press.
- 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 R, 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 MA, 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, Flenning JM, Myers RA, 2005. Robust state-space modeling of animal movement data. *Ecology* **86**(11) : 2874–2880.
- 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.
- 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 RECOPECA 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*, volume 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 J, Haydon D, Frair J, Holsiner K, Fryxell J, 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 Fisheries and Aquatic Sciences* **57**(1) : 51–65.

- Poos JJ, 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. In *Proceedings of the IEEE*, 257–286.
- Shumway R, Stoffer D, 2000. *Time Series Analysis and Its Applications*. Springer Texts In Statistics, Springer-Verlag GmbH.
- 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.
- 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.
- 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.
- Zucchini W, MacDonald IL, 2009. *Hidden Markov Models for Time Series : An Introduction Using R*. CRC Press.

FIGURES

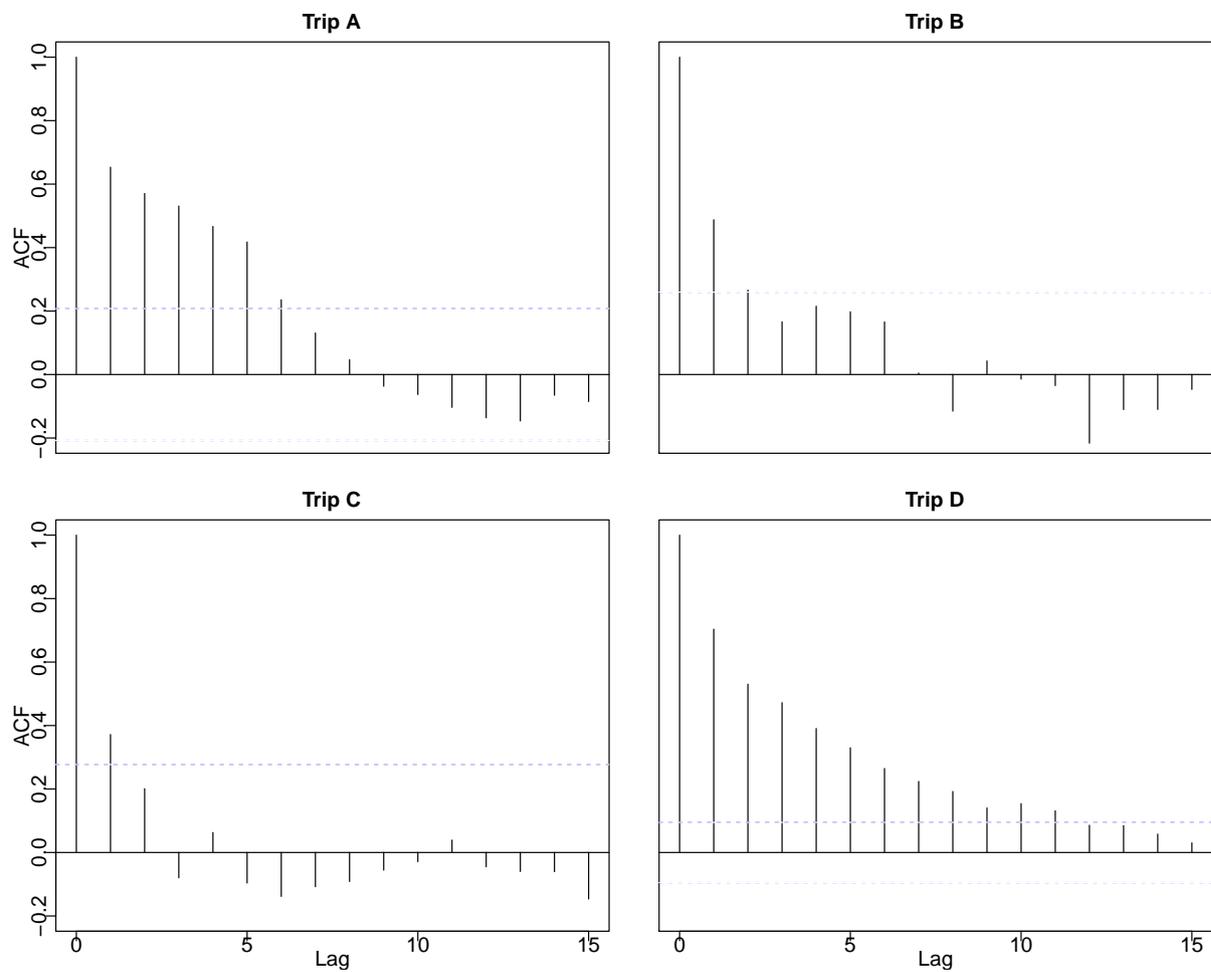


Figure 1. Autocorrelation function plots for trips A to D.

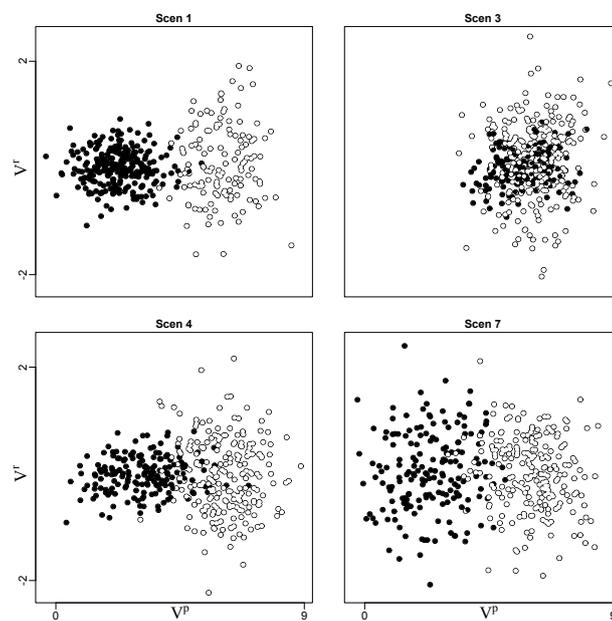


Figure 2. Simulated velocity processes for scenarios 1, 3, 4, and 7 (see Table 2). The persistent speed V^P is represented along the x axis and the turning speed is represented along the y axis.

. Black dots are for fishing, white dots for steaming.

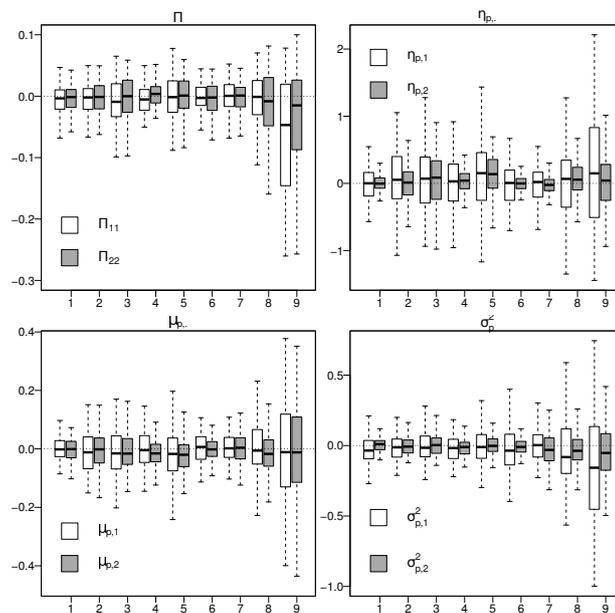


Figure 3. Box plots of estimation errors (estimated value minus the true value of each parameter, on the y axis) obtained for the 9 simulation scenarios (x axis) presented in Table 2. Only estimation errors for process V^p are presented, white and grey box plots are for parameters estimates in steaming and fishing respectively. The whiskers represent here at most 1.5 times the interquartile range. Outliers are not plotted.

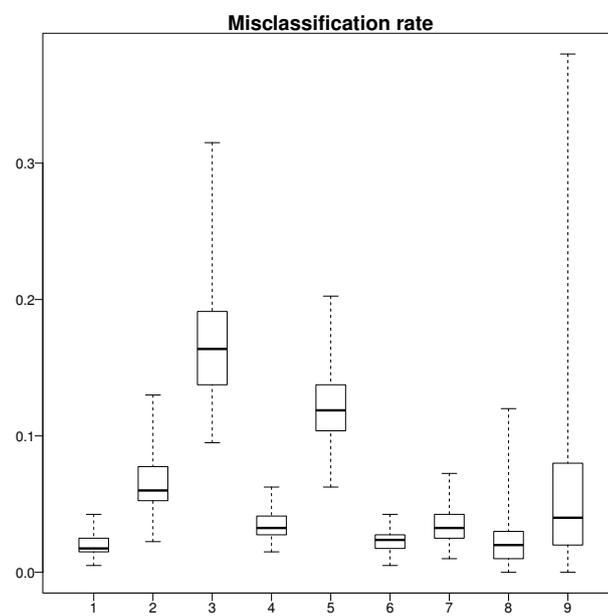


Figure 4. Box plot of misclassification rate from the Viterbi algorithm (on the y axis) for the 9 simulation scenarios (on the x axis) presented in Table 2, the whiskers height is at most 2 times the interquartile range. Outliers are not plotted.

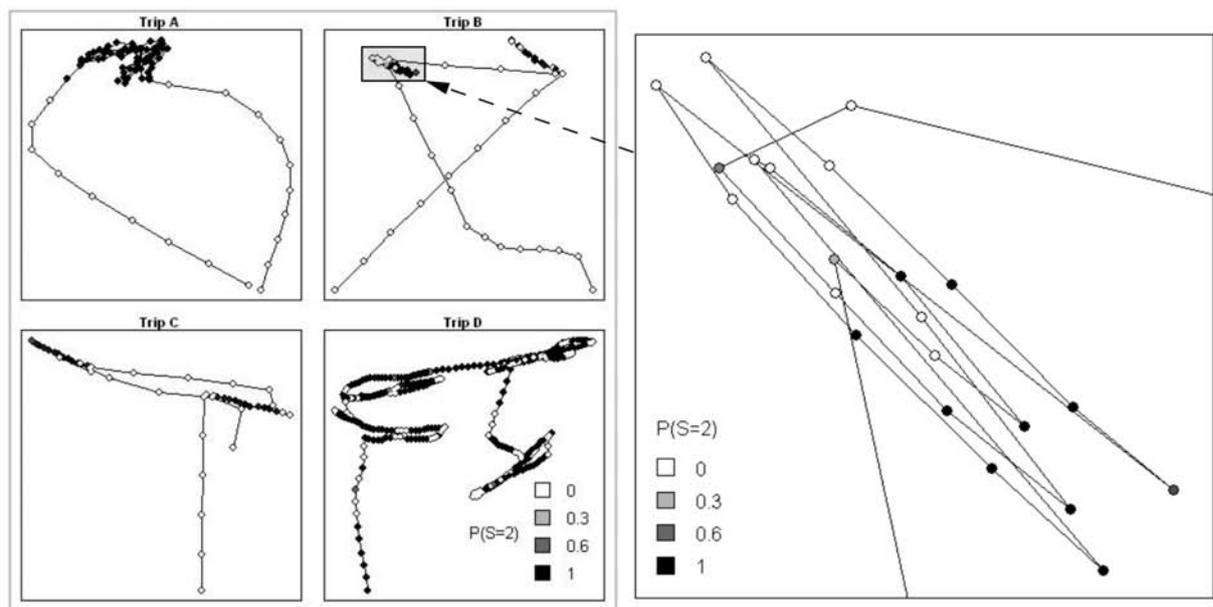


Figure 5. Estimated probabilities of being in state 2 (fishing) for each trip (A, B, C and D) is plotted, from 0 (white dots) to 1 (black dots). A zoom is made over a specific zone of trip B. Estimations for trip D can not be interpreted as steaming/fishing.

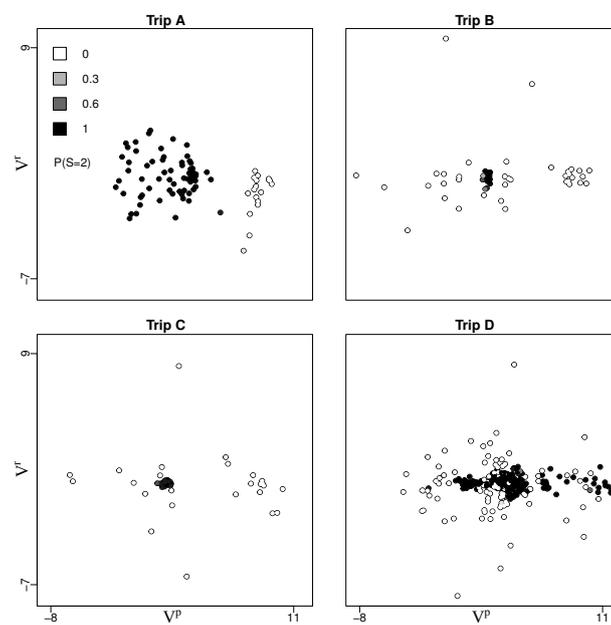


Figure 6. Velocity process for each trip (A, B, C and D). The persistent speed V^P is represented along the x axis and the turning speed is represented along the y axis.. The estimated probability of being in state 2 (fishing) for each trip is plotted, from 0 (white dots) to 1 (black dots).

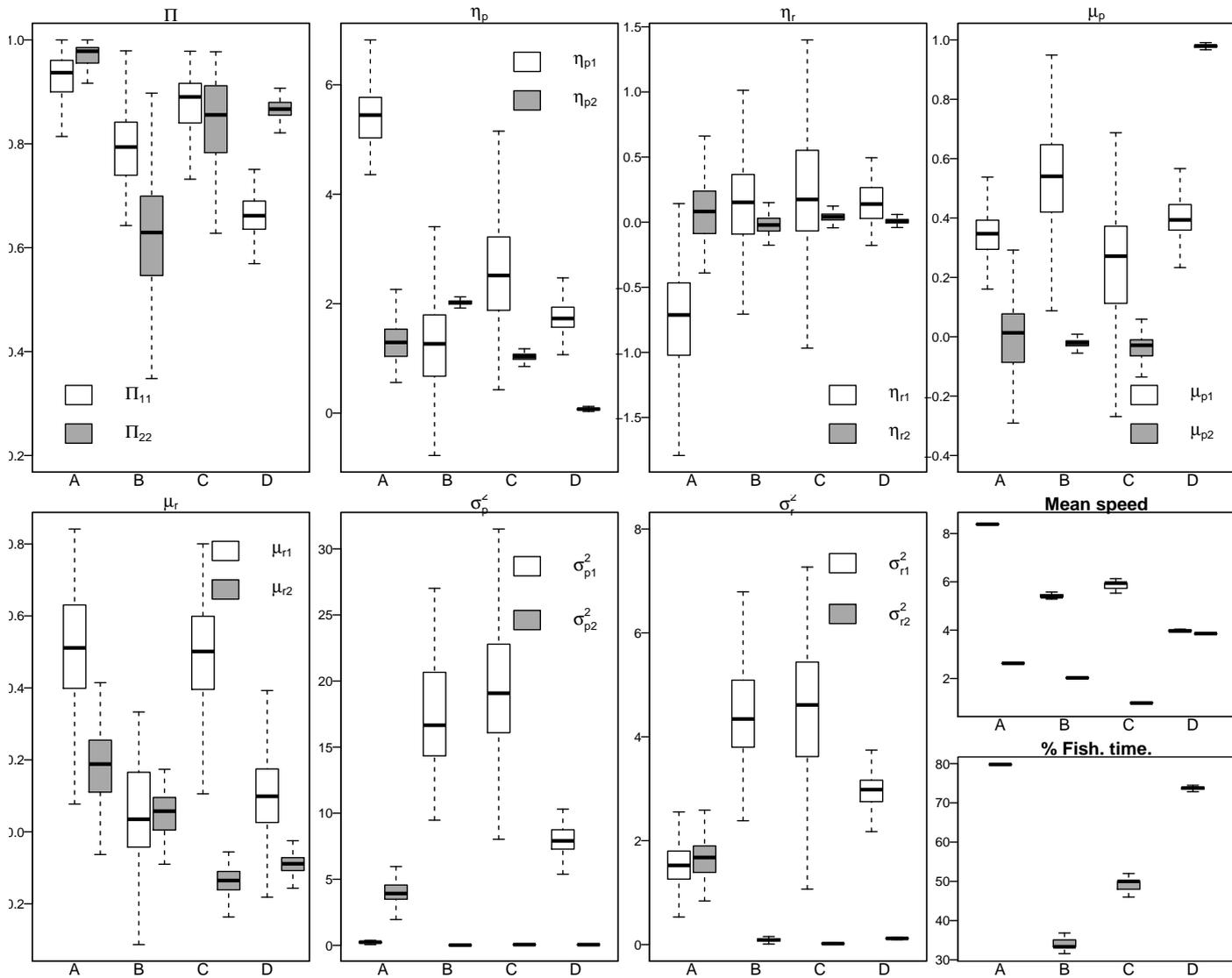


Figure 7. Box plots of estimations for parameters in each trip. Box plots of estimated mean speed in each state and estimated proportion of time spent fishing are also presented. White boxes stand for state 1, grey ones for state 2.

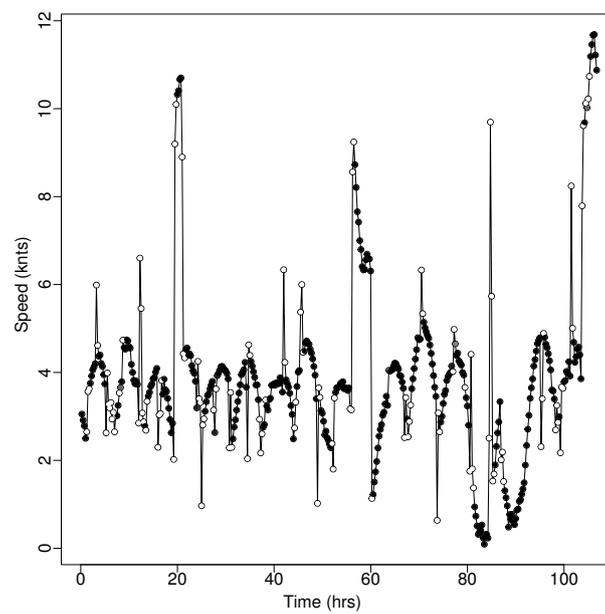


Figure 8. Scalar speed process for trip D, with estimated states. Probability of being in state 2 is plotted from 0 (white dots) to 1 (black dots). Scalar speed is high for both states. The state attribution here is based on autocorrelation (black dots and white dots for high and low autocorrelation,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

TABLES

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

Table 1. Technical details of the four studied trajectories

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

Table 2. Parameters values for each simulation scenario. The matrix Π is identical for all scenarios. Expectation and variance indicated are calculated from equations (5) and (6), and rounded to 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.