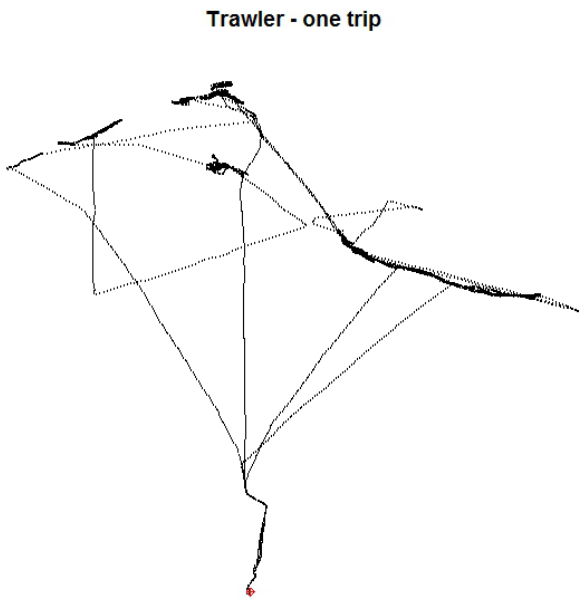
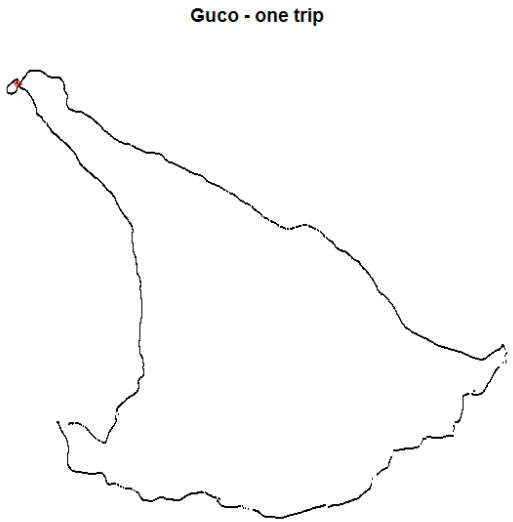
Validation data : keystone to assess performance of state-space models for movement

Stéphanie Mahévas (Ifremer), Nicolas Bez (IRD), Marie-Pierre Etienne (AgroParisTech),Pascal Monestiez (INRA), Etienne Rivot (AgrocampusOuest), Pierre Gloaguen (Ifremer), Youen Vermard (Ifremer), Mathieu Woillez (Ifremer), Sophie Bertrand (IRD), Rocio Joo (IMARPE), Maud Delattre (AgroParisTech), David Nerini (MIO), Emilie Walker (INRA), Hélène de Pontual (Ifremer)

3 juillet 2014, ISEC Montpellier

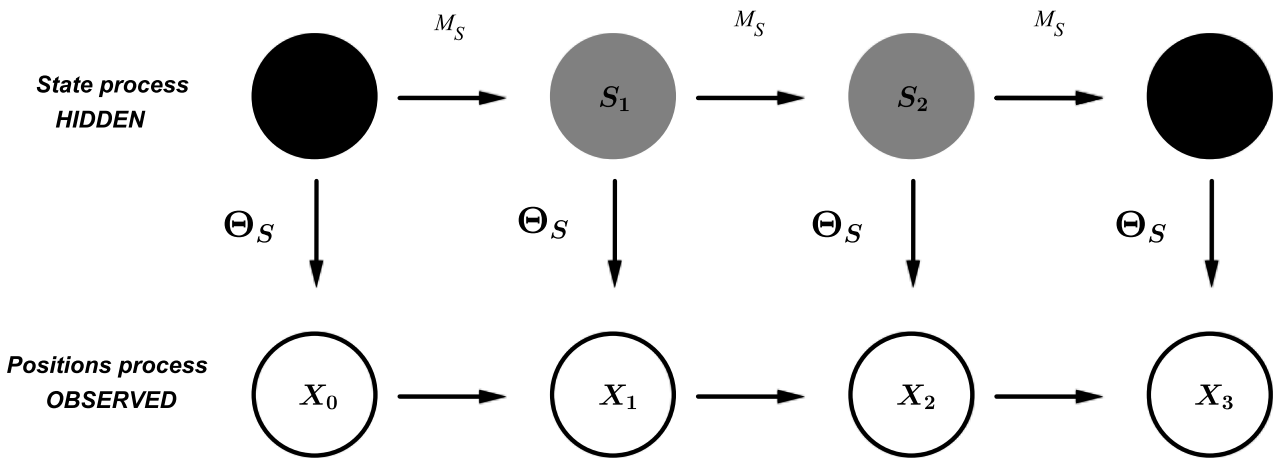
Inferring behavior from tracking data



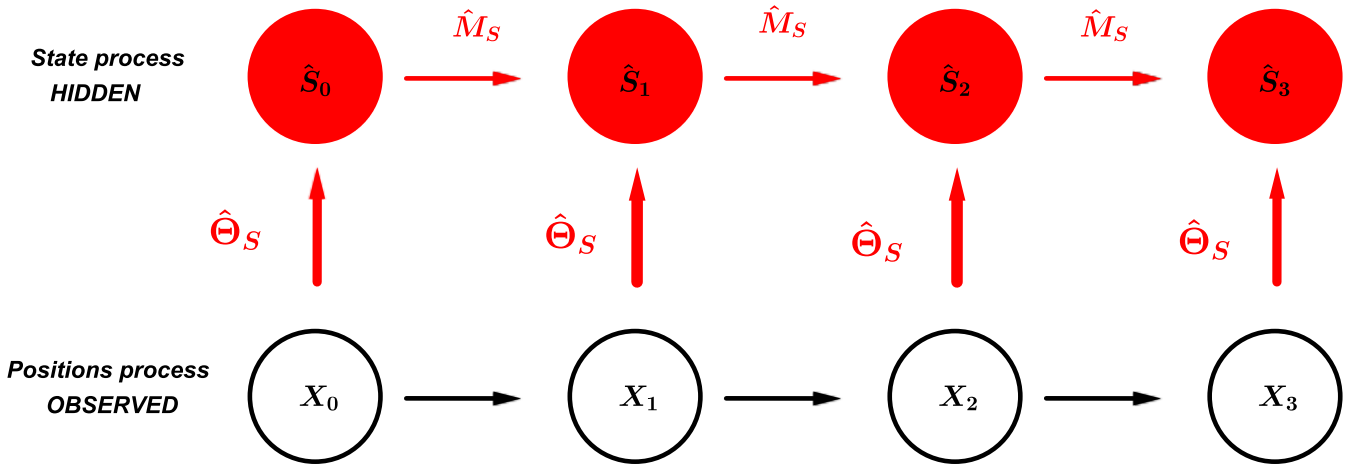


- Observations : positions at regular time step
- Behavior : fishing / not fishing, diving / not diving
- Learning individual behavior along trajectory from movement

Selecting a state space model



Selecting a state space model



Estimation : $\hat{\Theta}_S, \hat{M}_S$ and hidden states sequence$(\hat{S}_i))_i$

...required assumptions

1. Observed positions process $X_{t+1} = D_t + X_t + \epsilon_t$

assuming a piecewise linear path with

Uncorrelated process : V **velocity and** Ψ **turning angle**

$$V_t|(S_t) = f_V(\Theta_V) \text{ and } \Psi_t|(S_t) = f_\Psi(\Theta_\Psi)$$

(Vermard et al 2010, Walker and Bez 2010, Joo et al 2013)

Correlated process : **Speed** $V^p = Vcos(\Psi)$ **and** $V^r = Vsin(\Psi)$

$$V^p_{t+1}|(S_{t+1} = i) = \eta_{p,i} + \mu_{p,i} V^p_t + \sigma_{p,i}\epsilon_{p,t}$$
$$V^r_{t+1}|(S_{t+1} = i) = \eta_{r,i} + \mu_{r,i} V^r_t + \sigma_{r,i}\epsilon_{r,t}$$

(Gloaguen et al 2014)

2. Hidden states process $S_{t+1} = F((S_1, ..., S_t), M_S)$

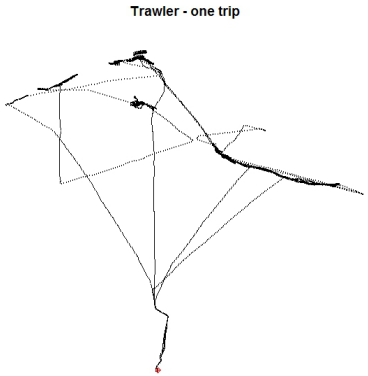
Markov Chain : $S_{t+1} = F(S_t, M_S)$

Semi Markov : $(S_{t_k}, T_{t_k})_{t_k}$ and $\tau_{t_k} = T_{t_k+1} - T_{t_k}$

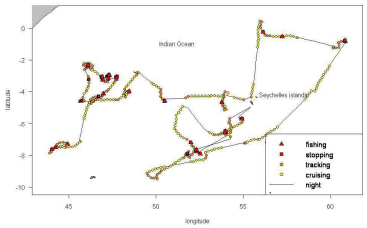
$$(S_{t_k+1}, \tau_{t_k+1}) = F((S_{t_1}, T_{t_1}, ..., S_{t_k}, T_{t_k}), M_S) = F(S_{t_k}, M_S)$$

High resolution trajectories with validation data

- Vessels and birds paths monitored **regularly** (with a **smaller time step** than usual)
- At each position, the state (fishing or not for vessels, diving or not for birds) is observed



SIH



High resolution trajectories with validation data

Learning from validation data

- are models assumptions violated ?
- are inferred behaviors robust to model assumptions ?
- are the answers sensitive to observations time step δ_t ?

Proposed approach

Degrading the observations time step,

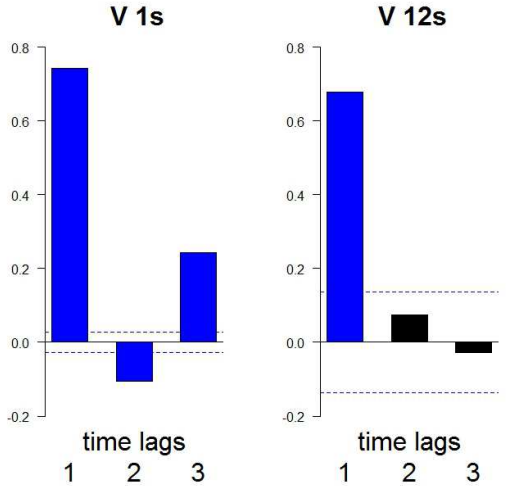
1. we explore
   - Positions process : is it correlated ? PACF analyses (H_0 : uncorrelated)
   - State process : is it Markov ? Residence time analyses (H_0 : geometric distribution)
2. we fit state space models with different assumptions
   - models' performance to predict hidden state sequence
   - are performances sensitive to assumptions ?

Exploration of model assumptions

1.Positions process :

$H_0 = V$ uncorrelated

Partial autocorrelation of V at lag 1, 2 and 3 for several δ_t



Autocorrelated ?

V, V_p, V_r 1rst(or 2nd) order correlated $\forall\ \delta_t$ and S

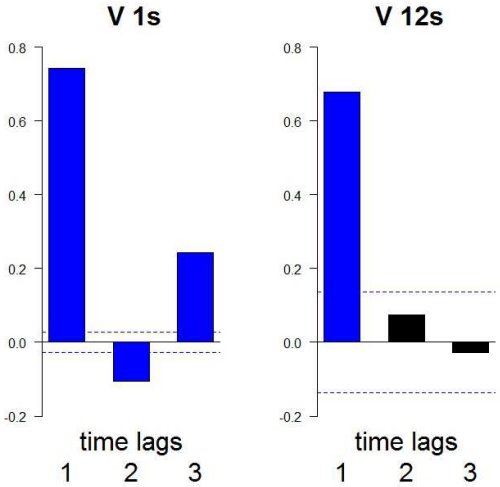
$\Rightarrow$ **autocorrolated process more appropriate**

Exploration of model assumptions

1.Positions process :

$H_0 = V$ uncorrelated

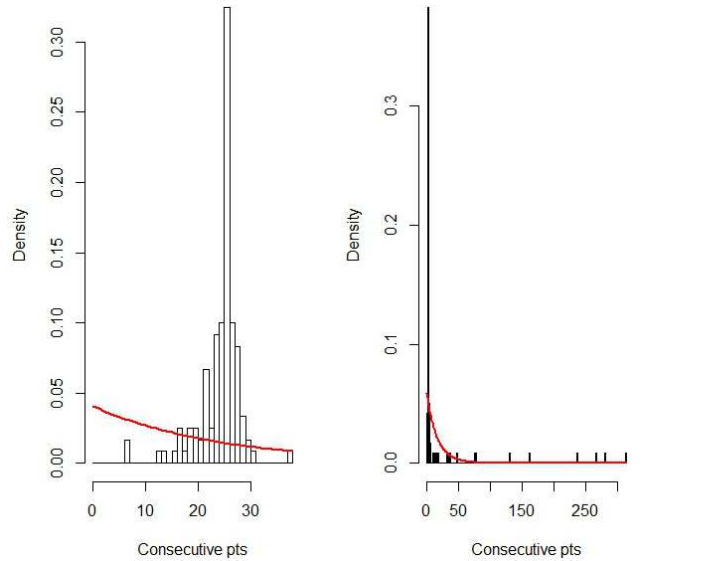
Partial autocorrelation of V at lag 1, 2 and 3 for several δ_t



2. State process :

H_0 = Residence time $\sim$ geometric

test de χ_2 : geometric(p = empiric mean)



Autocorrelated ?

V, V_p, V_r 1rst(or 2nd) order correlated $\forall$ δ_t and S

$\Rightarrow$ **autocorrolated process more appropriate**

Markov ?

Only relevant for Not Fishing or Not Diving $\Rightarrow$ Semi-Markov more appropriate

Fitting State Space Models (fisheries)

		Position	
	Assumptions	Uncorrelated : (V, Ψ)	Autoregressive : (V_p, V_r)
State	Markov	X	X
	Semi-Markov	X	X

Steps of the experiment for each SPM

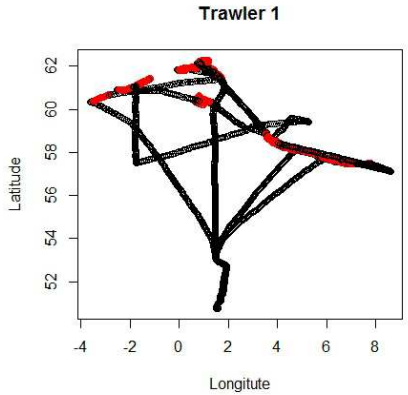
1. Split the dataset into a learning dataset and a testing dataset
2. With the **learning trajectories** : estimate the model parameters
   - speed and turning angle distributions, correlations
   - residence time for transition matrix
3. With the **remaining trajectories** :
   - simulating the most likely sequence of states using the Viterbi algorithm
   - estimating the performance of the model : confusion matrix

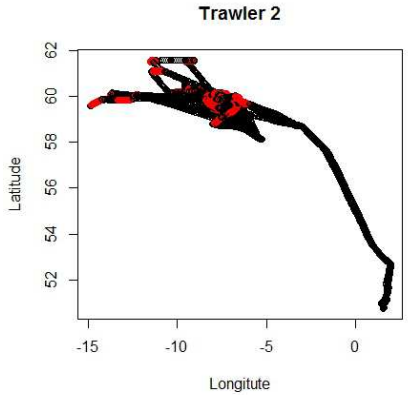
The most likely sequence of states

Uncorrelated : (V, Ψ)

Using **learning trajectories** for two trawlers (5 and 13 trips)

Trawler 1 and Trawler 2



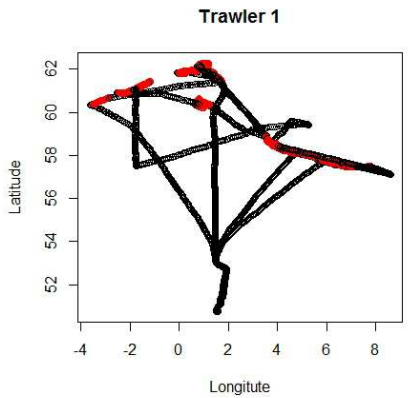


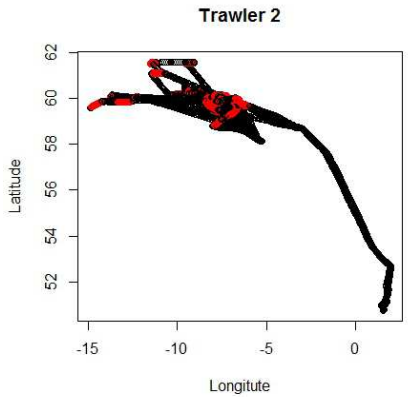
The most likely sequence of states

Uncorrelated : (V, Ψ)

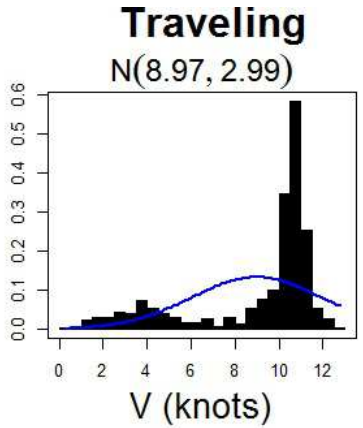
Using **learning trajectories** for two trawlers (5 and 13 trips)

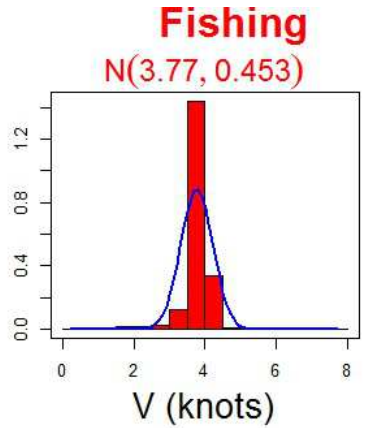
Trawler 1 and Trawler 2

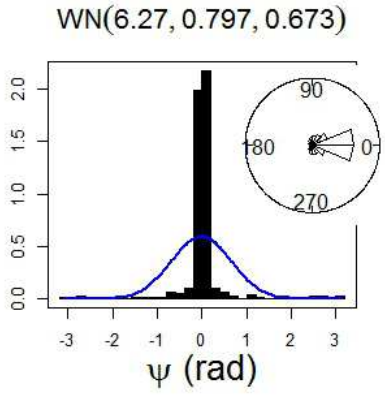


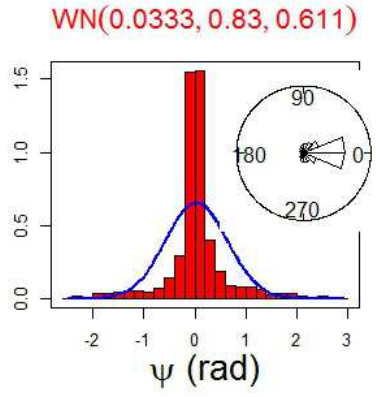


Trawler 1 with $\delta_t = 15min$: $\hat{}_V$ and $\hat{}$-









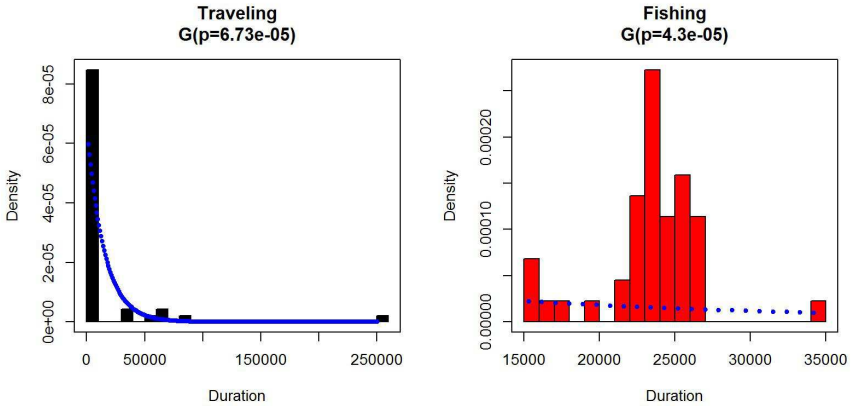
The most likely sequence of states

Uncorrelated : (V, Ψ)

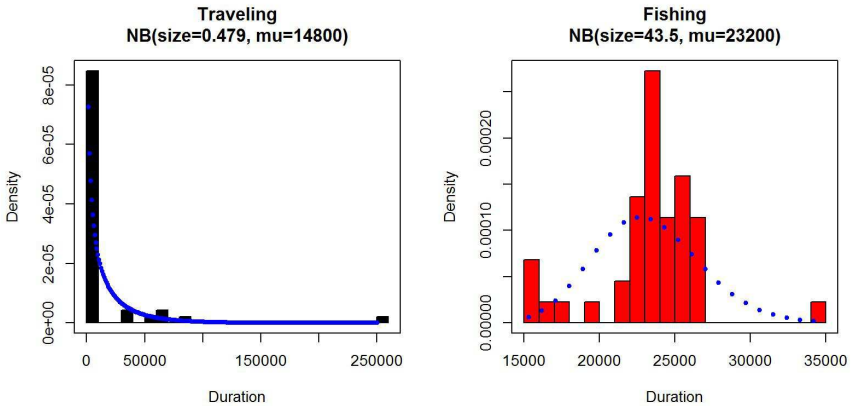
Using **learning trajectories**

Trawler 1 with $\delta_t = 15min$: $\hat{M}_S$

Markov



Semi Markov



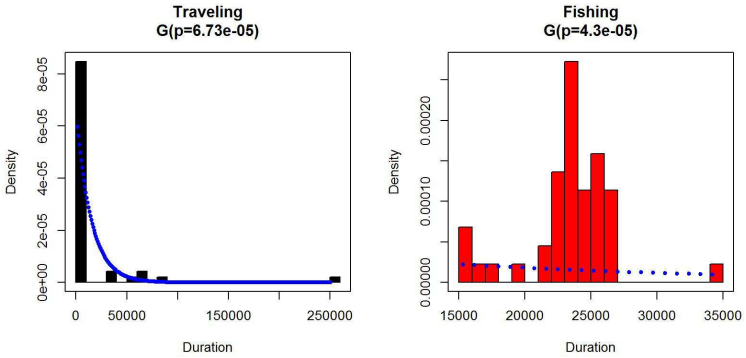
The most likely sequence of states

Uncorrelated : (V, Ψ)

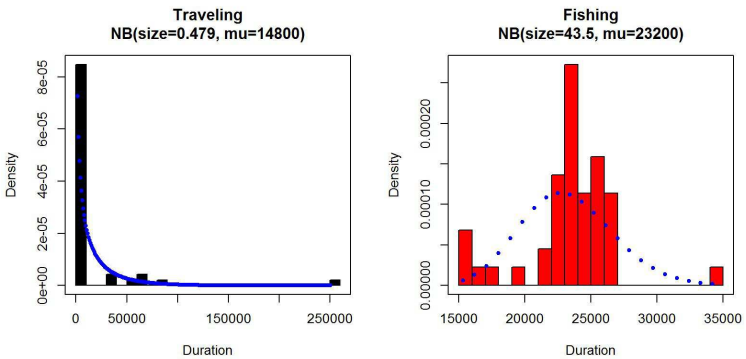
Using **learning trajectories**

Trawler 1 with $\delta_t = 15min$: $\hat{M}_S$

Markov



Semi Markov



$\hat{\Theta}_V$, $\hat{\Theta}_\Psi$ and $\hat{M}_S$

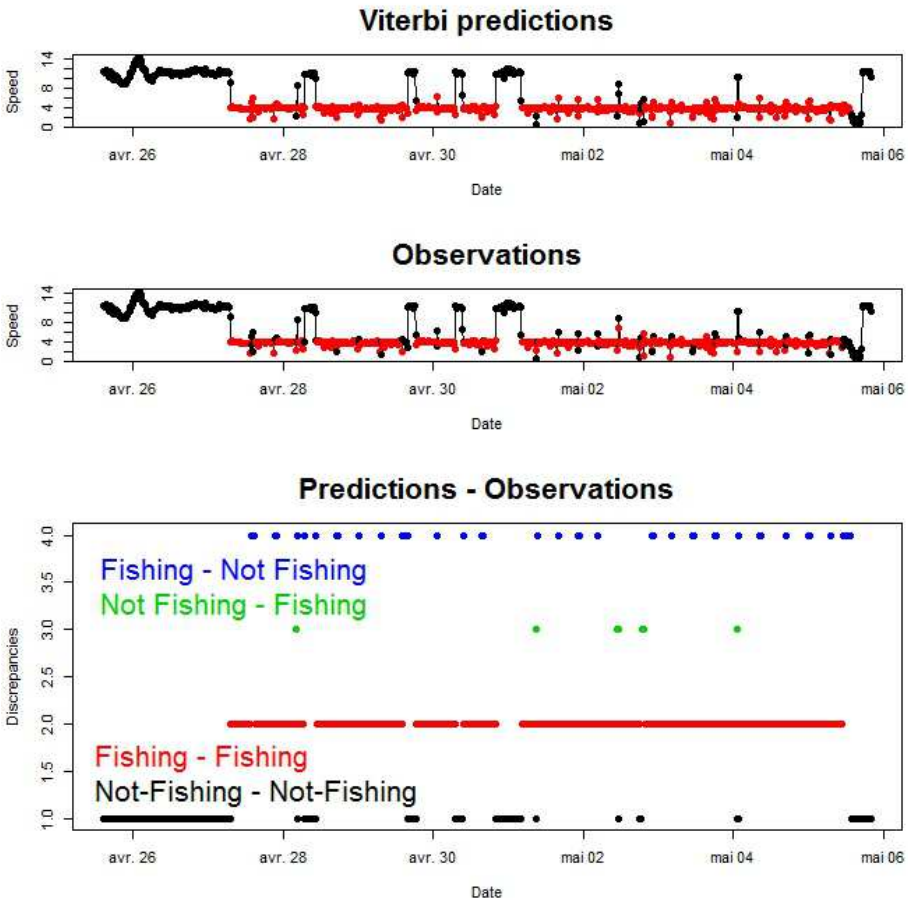
Statistical distributions do not fit well observed distributions of speeds, turning angles and residence times in fishing and not fishing states

The most likely sequence of states

Uncorrelated, (V, Ψ)

Using the **remaining trajectories** (Viterbi algorithm with $\hat{\Theta}_V$, $\hat{\Theta}_\Psi$ and $\hat{M}_S$)

Markov

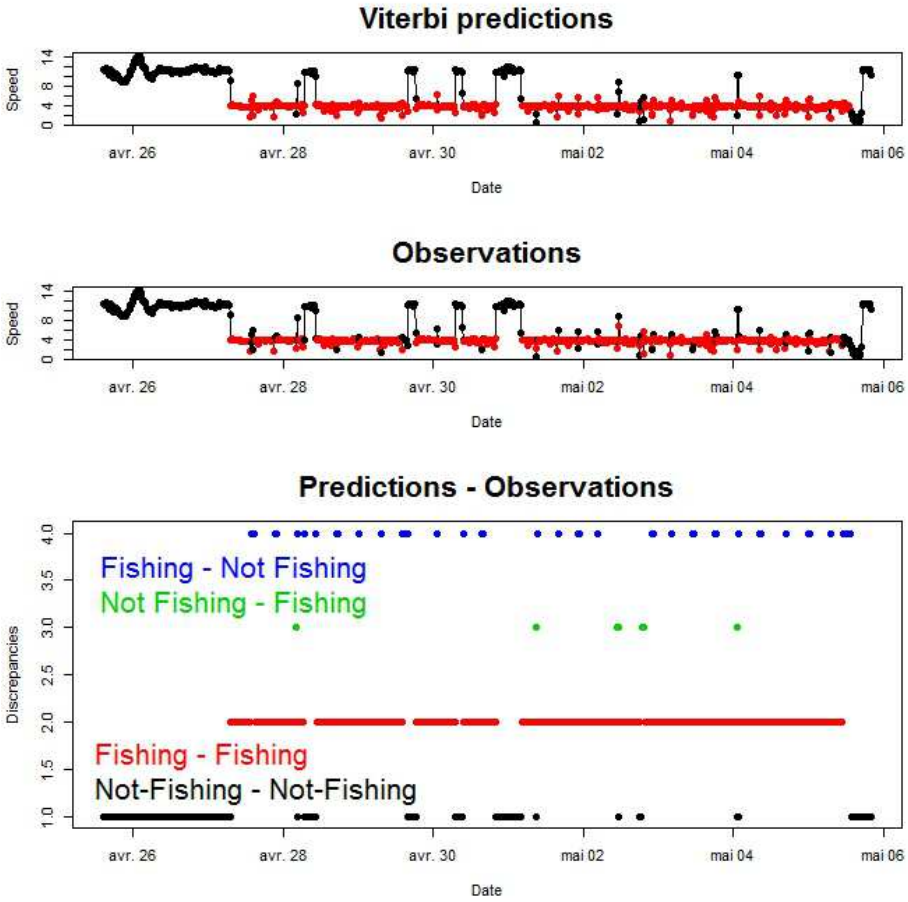


The most likely sequence of states

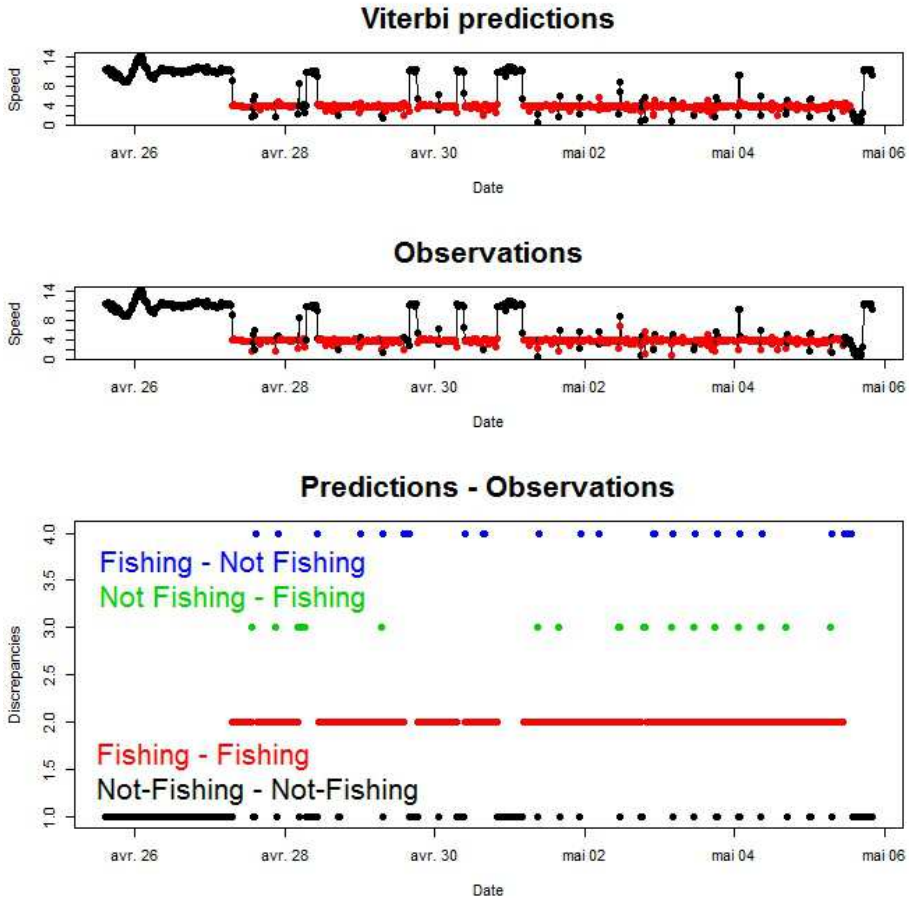
Uncorrelated, (V, Ψ)

Using the **remaining trajectories** (Viterbi algorithm with $\hat{\Theta}_V$, $\hat{\Theta}_\Psi$ and $\hat{M}_S$)

Markov



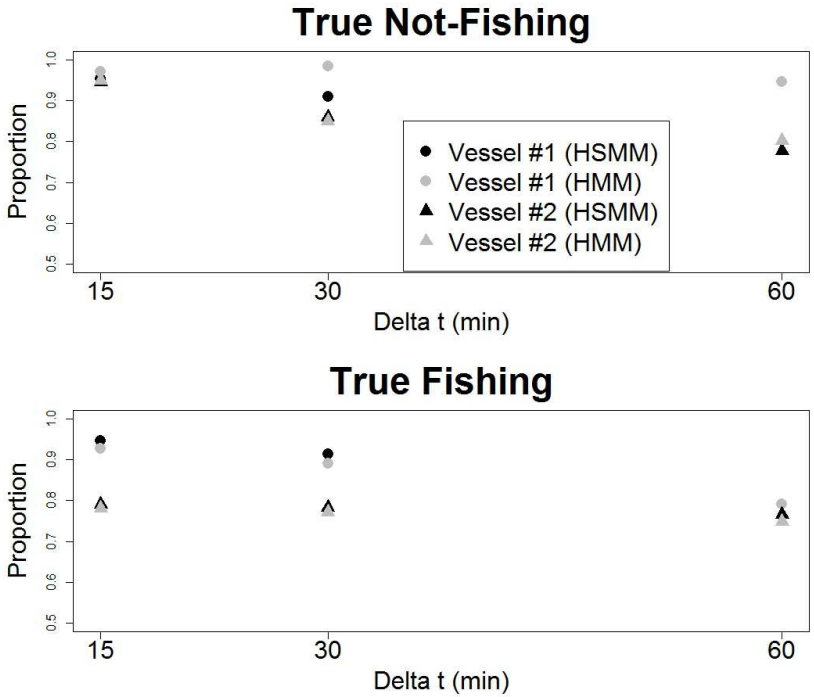
Semi Markov



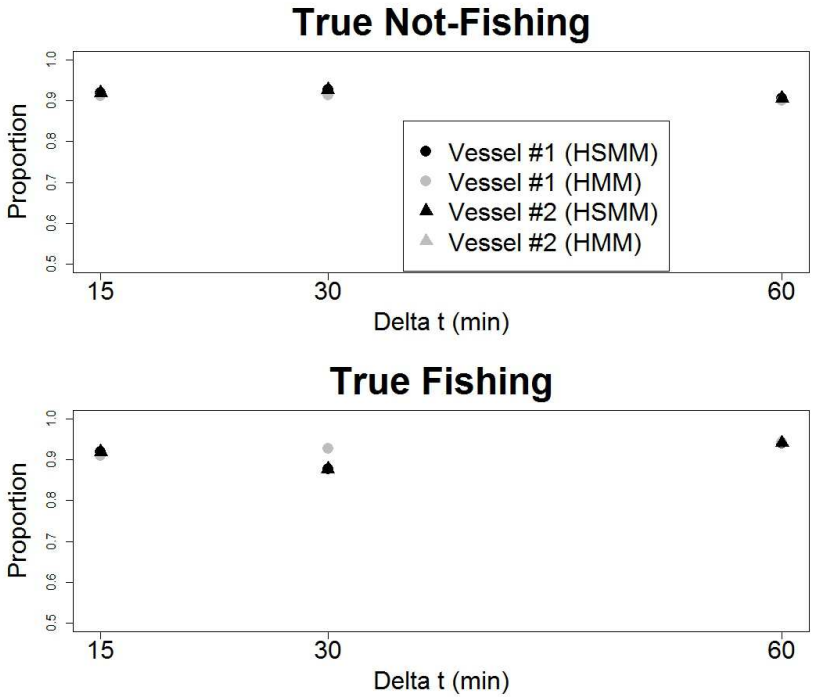
The most likely sequence of states - summary of models'performance

Confusion matrix

Uncorrelated Model



Autocorrelated Model

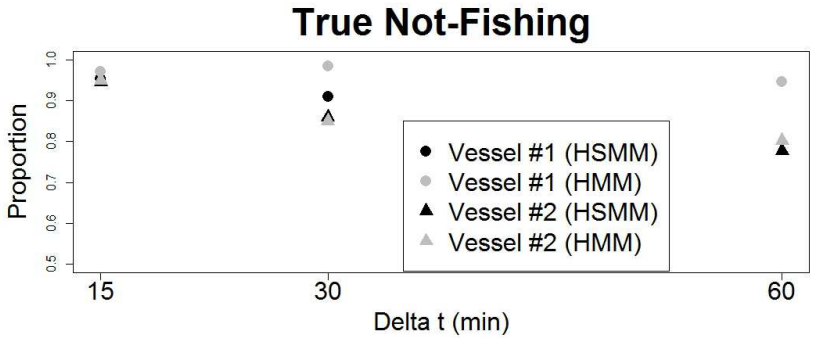


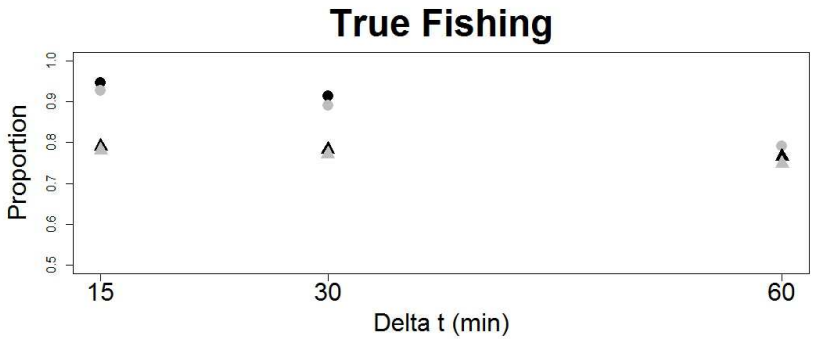
- **Proportions of true predictions** $\in [0.85, 0.95]$ whatever δ_t, behavior and positions processes assumptions ;
- **Best fit for the auto-correlated model with semi-Markov states transition**
- Small degradation with δ_t

The most likely sequence of states - summary of models'performance

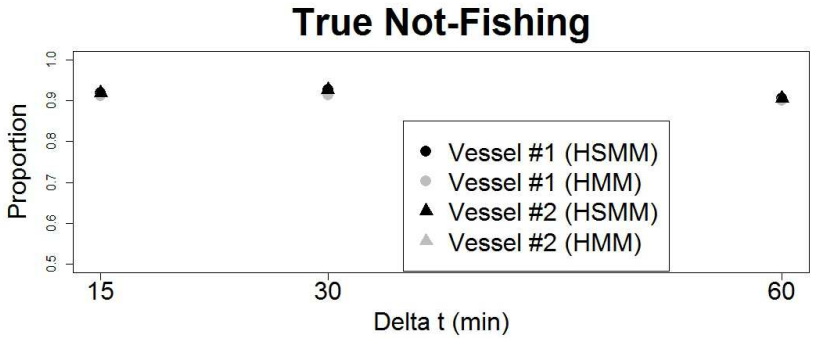
Confusion matrix

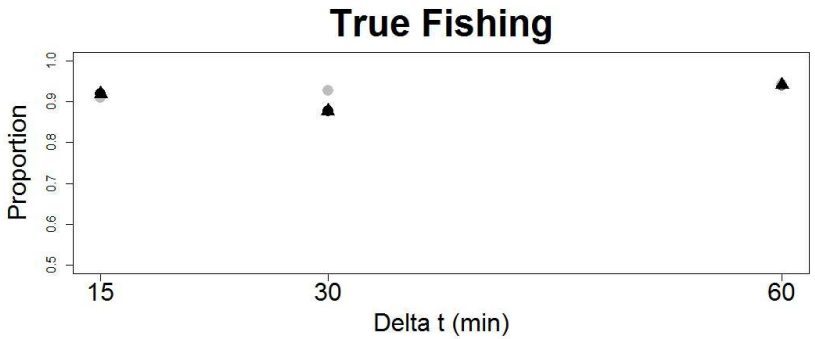
Uncorrelated Model





Autocorrelated Model





- **Proportions of true predictions** $\in [0.85, 0.95]$ whatever δ_t, behavior and positions processes assumptions ;
- **Best fit for the auto-correlated model with semi-Markov states transition**
- Small degradation with δ_t

Conclusions

- Are models assumptions violated ?
  - fisrt order correlation rarely taken into account (fisheries)
  - Markov only confirmed for not fishing or not diving state
- Are inferred behaviors robust to model assumptions ?
  - Uncorrelated and AR models are robust to state process assumptions
  - but fitted Θ distributions not satisfactory
- Are the answers sensitive to observations time step ?
  - small degradation of models'performance

perspectives

- Models'fit : same results for birds ?
- What is the influence of length trajectories on conclusions ?
- What is the influence of the learning step on conclusions ?

Conclusions

- Are models assumptions violated ?
  - fisrt order correlation rarely taken into account (fisheries)
  - Markov only confirmed for not fishing or not diving state
- Are inferred behaviors robust to model assumptions ?
  - Uncorrelated and AR models are robust to state process assumptions
  - but fitted Θ distributions not satisfactory
- Are the answers sensitive to observations time step ?
  - small degradation of models'performance

perspectives

- Models'fit : same results for birds ?
- What is the influence of length trajectories on conclusions ?
- What is the influence of the learning step on conclusions ?