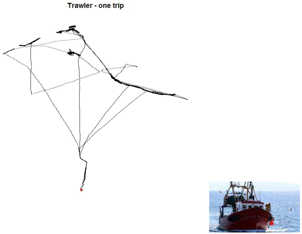
Validation data : keystone to assess performance of state-space models for movement

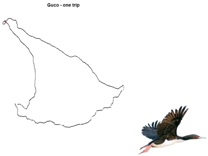
Stéphanie Mahévas (Ifremer), Nicolas Bez (IRD), Marie-Pierre Etienne (AgroParisTech),Pascal Monestiez (INRA), Etienne Rivot (AgrocampusOuest), Pierre Gloaguen (Ifremer), Youen Vermard (Ifremer), Mathieu Woillez (Ifremer), Sophie Bertrand (IRD), Rocio Joo (IMARPE), Maud Delattre (AgroParisTech), David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

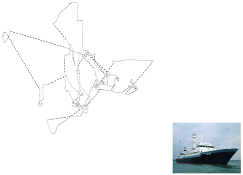
3 juillet 2014, ISEC Montpelier

Plan

Inferring behavior from tracking data

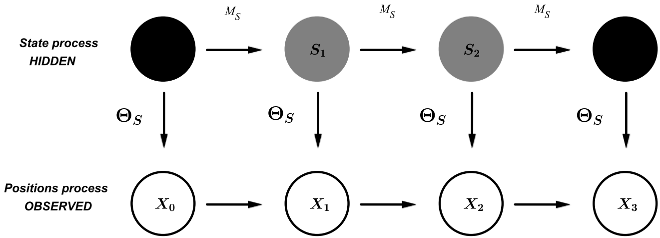




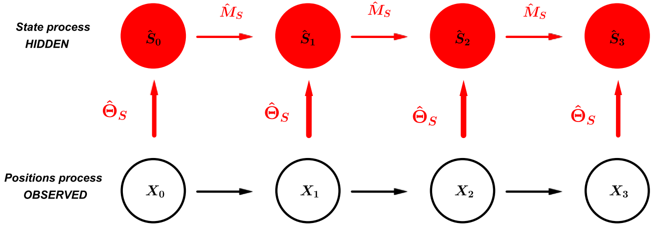


- Learning individual behavior along trajectory from movement
- Observations : positions at regular time step
- Behavior : fishing / not fishing, diving / not diving

Selecting a state space model

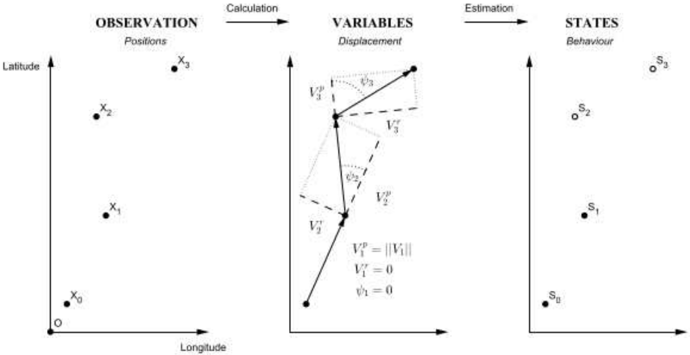


Selecting a state space model



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

Selecting a state space model



...required assumptions

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

assuming a piecewise linear path with

Uncorrelated process : $V_t|(S_t) = f_V(\Theta_V)$ and $\Psi_t|(S_t) = f_\Psi(\Theta_\Psi)$

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

Correlated process : $V^p_{t+1}|(S_{t+1} = i) = \eta_{p,i} + \mu_{p,i} V^p_t + \sigma_{p,i}\epsilon_{p,t}$

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

Autoregressive Model (Gloaguen et al 2014)

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

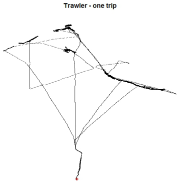
Markov Chain : $S_{t+1} = F(S_t, M_S)$

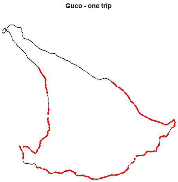
Semi Markov : $(S_{t_k}, T_{t_k})_{t_k}$ and $\tau_{t_k} = T_{t_k+1} - T_{t_k}$

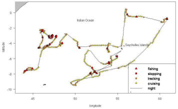
$$(S_{t_k+1}, \tau_{t_k+1}) = F((S_{t_1}, T_{t_1}, ..., S_{t_k}, T_{t_k}), M_S) = F(S_{t_k}, M_S)$$

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







High resolution trajectories with validation data

Learning from data and models fitting

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

Performance of state space models : experiment

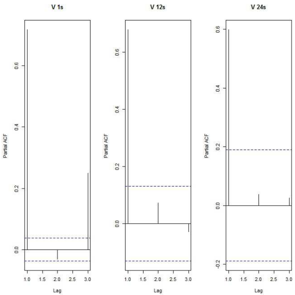
Degrading the observations time step,

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 estimate model parameters independantly (maximum likelihood)
3. we predict the most likely sequence of states using Viterbi algorithm

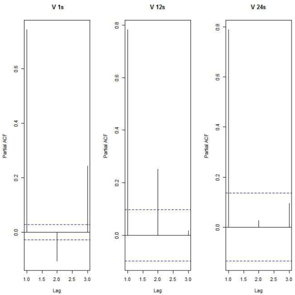
Positions process : $H_0 = V$ uncorrelated

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

Diving



Not Diving



Autocorrelated ?

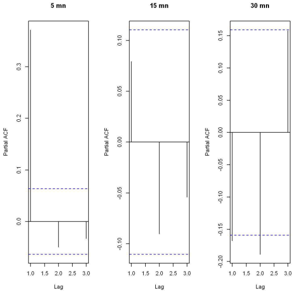
Diving, Not Diving : V, V_p, V_r first(or second) order correlated

δ_t : V, V_p, V_r first order correlated whatever δ_t

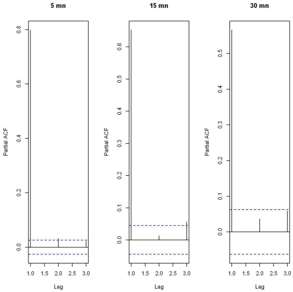
Positions process : $H_0 = V$ uncorrelated

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

Fishing



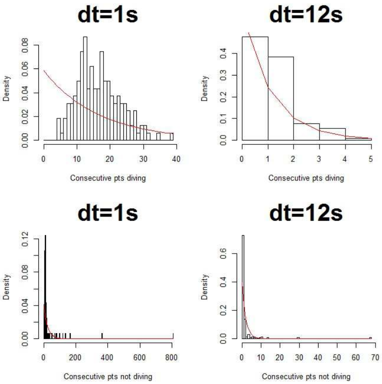
Not Fishing



Autocorrelated ?

Fishing, Not Fishing : V, V_p, V_r first (or second) order correlation whatever $\delta_t = 15min, 1h$

Behavior process : H_0 = Residence time $\sim$ geometric distribution



δ_t	pval fish	pval steam
1s	0.01	0.28
12s	0.000	0.000

test de χ_2 : geometric (p = empiric mean)

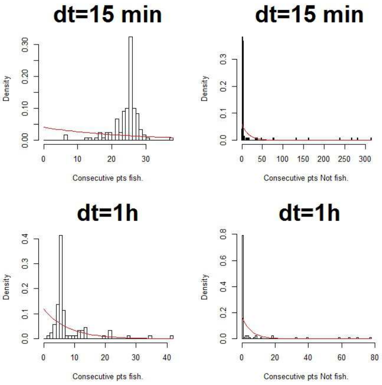
number of consecutive positions

Markov ?

Diving : not relevant - better should be semi-Markov

Not Diving : Markov

Behavior process : H_0 = Residence time $\sim$ geometric distribution



δ_t	pval fish	pval steam
15min	0.002	0.000
1h	0.000	0.000

test de χ_2 : geometric (p = empiric mean)

number of consecutive positions

Markov ?

Fishing : not relevant - better should be semi-Markov

Not Fishing : Markov

Fisheries Modelling experiment

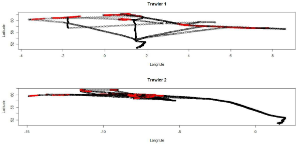
Assumptions	Uncorrelated, (V, Ψ)	Autoregressive, (V_p, V_r)
Markov	X	X
Semi-Markov	X	X

Steps of the experiment

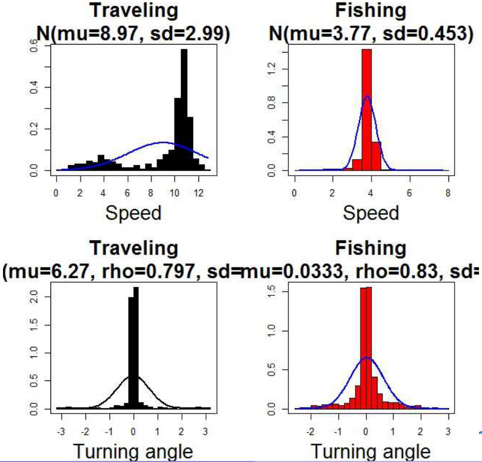
1. Split the dataset into a learning dataset and a testing dataset
2. With the learning dataset : estimate the model parameters
   - speed and turning angle distributions, correlations
   - residence time for transition matrix
3. With the remaining dataset :
   - simulating the most likely sequence of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

The most likely sequence of states- Uncorrelated, (V, Ψ)

Analysis performed using learning dataset for two trawlers operating in the English Channel (5 and 13 trips)



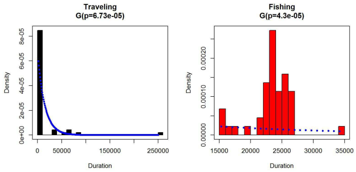
For trawler 1 with $\delta_t = 15min$ estimates of Θ_V and Θ_Ψ



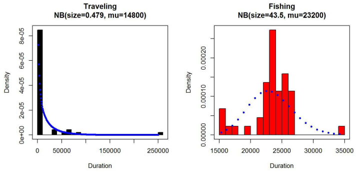
The most likely sequence of states - Uncorrelated, (V, Ψ)

For trawler 1 with $\delta_t = 15min$ estimates of M_S

Markov



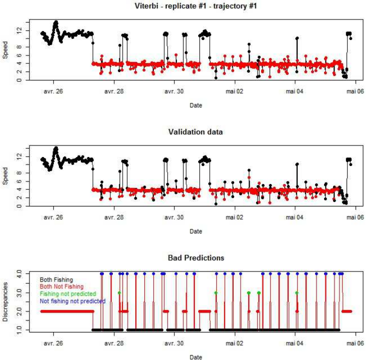
Semi Markov



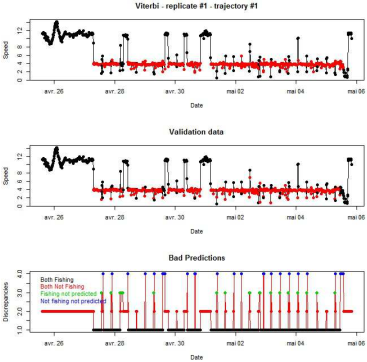
The most likely sequence of states - Uncorrelated, (V, Ψ)

Using the Viterbi algorithm with the estimated parameters Θ_V, Θ_Ψ and M_S

Markov

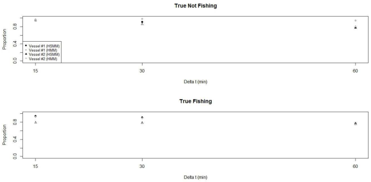


Semi Markov

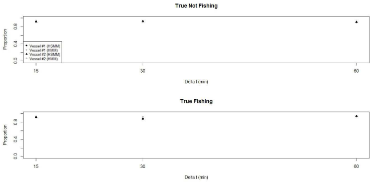


The most likely sequence of states - summary of the whole experiment

Uncorrelated Model



Autocorrelated Model



Performance of AR model and Drift model

Best fit for vessel 1

Robust to Markov assumption

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 ?
  - increasing time step increases autocorrelation for birds'speed
  - increasing time step decreases autocorrelation for vessels'speed
  - but does not influence the performance of AR and uncorrelated model

Conclusions

- Ecology - Fisheries : differences ?
- Time step : a limitation for fishing sequence identification but not for inferring behavior
- Speed is not observed - computing average speed required assumptions on the path between two positions - monitor instantaneous speed
- AR model or Drift Model ? the simplest Drift-Markov model
- Apropriate approach for trajectories with validation data
- Next step : what is the influence of the learning step on conclusions ?
  - performance of the autocorrelated model are known to be lower than those of the uncorrelated
  - to be compared with an integrated estimation procedure (like with not observed behavior)
- another track : segmentation of path (using path chracteristics, patterns ...), continuous model