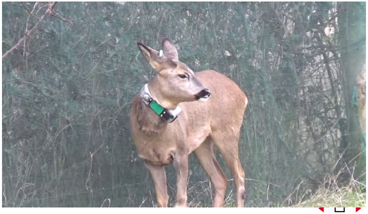
Supervised HSMM to study the behavioral sequences of wildlife in nature from accelerometer data.

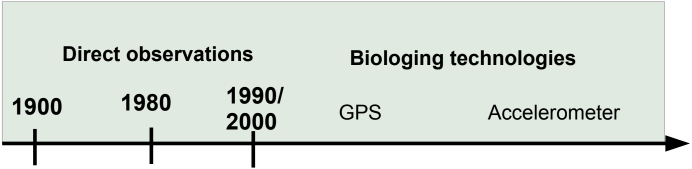
Sandra Plancade, Nathalie Peyrard
(Applied Mathematics and Informatics, INRAE, Toulouse)
Nicolas Morellet,Nathan Ranc
(Center of wildlife study, INRAE, Toulouse)

MASEMO - 1st-4th July 2025



Behavioral ecology

How animals behave to satisfy their fundamental needs



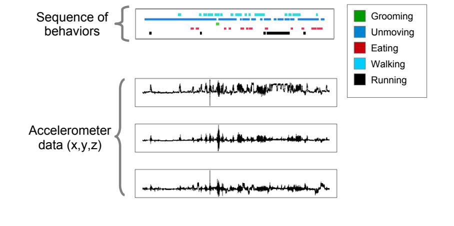



Amount of data	weak	good	very good
Precision of information on behavior	very good	gross	very fine scale but limited methods

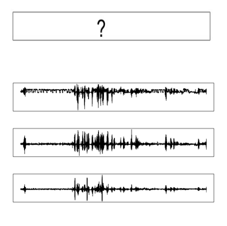
Supervised framework

- Captive animals : videos $\sim$ 10 minutes + accelerometer (32hz)
- Wild animals : accelerometer $\sim$ months

Observed animals



Unobserved animals

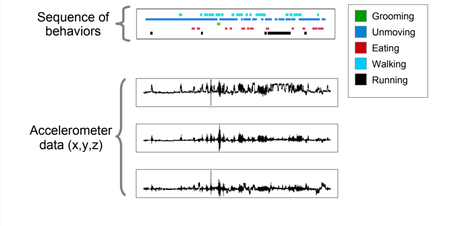


Supervised framework

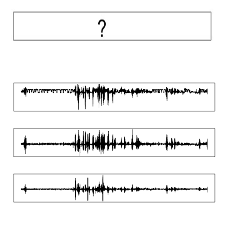
- Captive animals : videos $\sim$ 10 minutes + accelerometer (32hz)
- Wild animals : accelerometer $\sim$ months

$\hookrightarrow$ To evaluate the model : train/test set

Observed animals



Unobserved animals

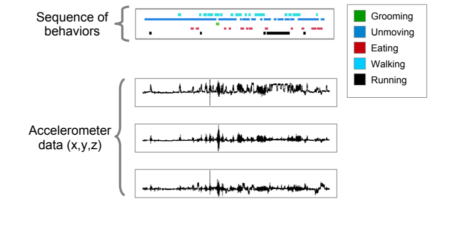


Supervised framework

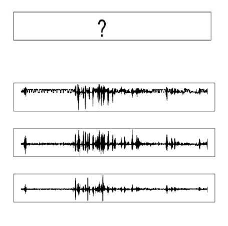
- Captive animals : videos $\sim$ 10 minutes + accelerometer (32hz)
- Wild animals : accelerometer $\sim$ months

$\hookrightarrow$ To evaluate the model : train/test set

Observed animals

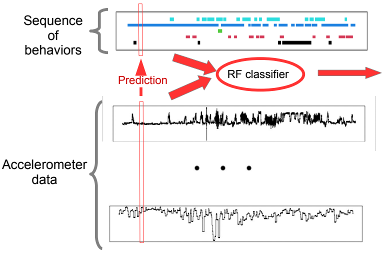


Unobserved animals



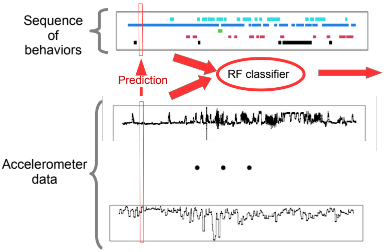
Classic method : Time-by-time classification

Observed animals

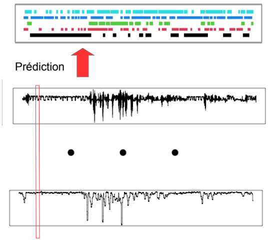


Classic method : Time-by-time classification

Observed animals

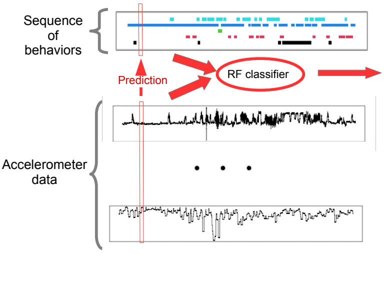


Unobserved animals

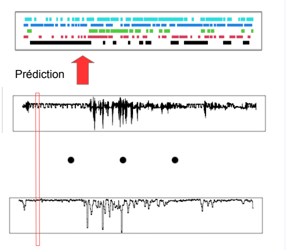


Classic method : Time-by-time classification

Observed animals

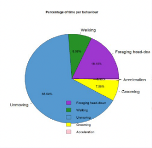


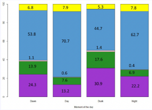
Unobserved animals



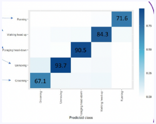
Output

Time budget

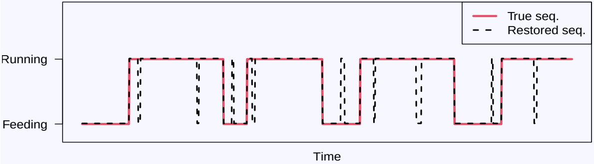




Model evaluation = accuracy

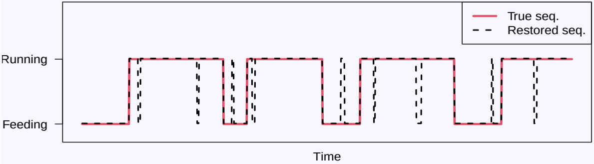


Good accuracy $\neq$ good sequence restoration



- Accuracy : 0.95
- Wrong transitions
- Wrong very short durations

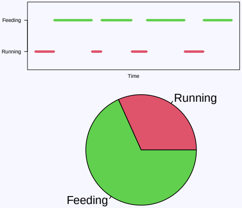
Good accuracy ≠ good sequence restoration



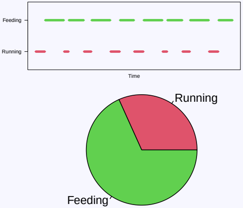
- Accuracy : 0.95
- Wrong transitions
- Wrong very short durations

Why sequence analysis?

Forest landscape

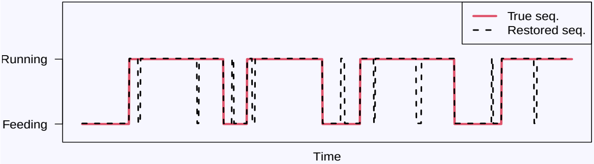


Open landscape



- Different dynamics
- Same time budget

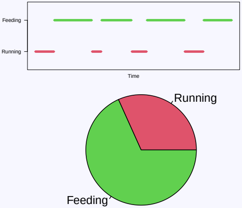
Good accuracy $\neq$ good sequence restoration



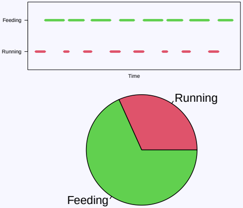
- Accuracy : 0.95
- Wrong transitions
- Wrong very short durations

Why sequence analysis?

Forest landscape



Open landscape

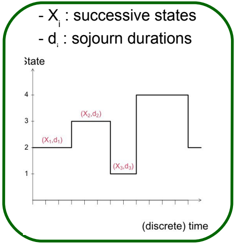


- Different dynamics
- Same time budget

$\Rightarrow$ HSMM : account for behavior durations and transitions

Hidden Semi-Markov Models

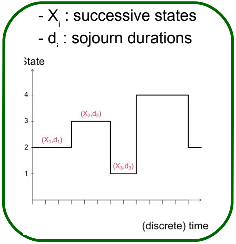
SMM



$$\mathbb{P}[(X_{i+1}, d_{i+1})|(X_i, d_i), \cancel{(X_{i-1}, d_{i-1})}, \ldots, \cancel{(X_1, d_1)}] \overset{\text{ED}}{=} \mathbb{P}[d_{i+1}|X_{i+1}]\mathbb{P}[X_{i+1}|X_i]$$

Hidden Semi-Markov Models

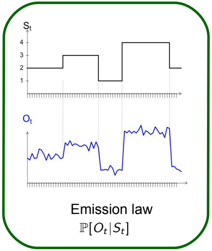
SMM



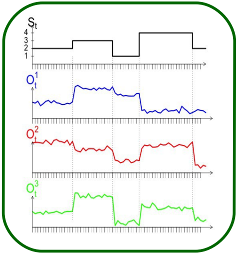
HSMM

Semi-Markov process

Observed process



multiresponse HSMM

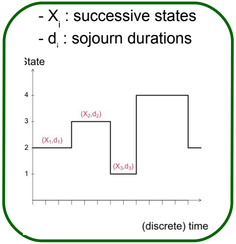


$$\mathbb{P}[(X_{i+1},d_{i+1})|(X_i,d_i),\cancel{(X_{i-1},d_{i-1})},\ldots,\cancel{(X_1,d_1)}] \overset{ED}{=} \mathbb{P}[d_{i+1}|X_{i+1}]\mathbb{P}[X_{i+1}|X_i]$$

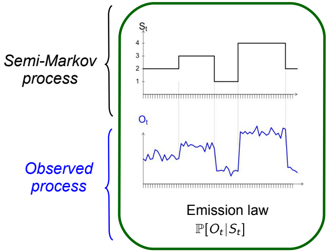
$$\mathbb{P}[(O_t)_t|(S_t)_t] = \prod_t \mathbb{P}[O_t|S_t]$$

Hidden Semi-Markov Models

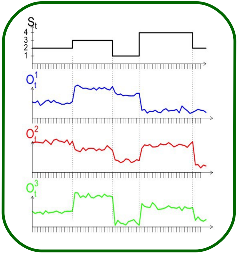
SMM



HSMM



multiresponse HSMM



$$\mathbb{P}[(X_{i+1},d_{i+1})|(X_i,d_i),\cancel{(X_{i-1},d_{i-1})},\dots,\cancel{(X_1,d_1)}] \overset{ED}{=} \mathbb{P}[d_{i+1}|X_{i+1}]\mathbb{P}[X_{i+1}|X_i]$$

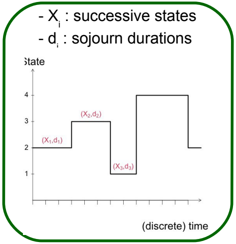
$$\mathbb{P}[(O_t)_t|(S_t)_t] = \prod_t \mathbb{P}[O_t|S_t]$$

Parameters (Explicit Duration)

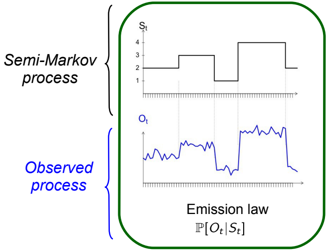
- Emission $\mathbb{P}_\Theta[O_t|S_t]$
- Sojourn duration $\mathbb{P}_\Theta[d_i|X_i]$
- Transition $\mathbb{P}_\Theta[E_{i+1}|X_i]$

Hidden Semi-Markov Models

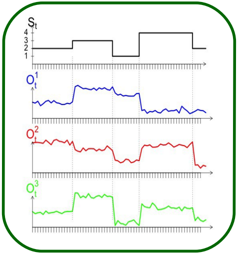
SMM



HSMM



multiresponse HSMM



$$\mathbb{P}[(X_{i+1},d_{i+1})|(X_i,d_i),\cancel{(X_{i-1},d_{i-1})},\dots,\cancel{(X_1,d_1)}] \overset{ED}{=} \mathbb{P}[d_{i+1}|X_{i+1}]\mathbb{P}[X_{i+1}|X_i]$$

$$\mathbb{P}[(O_t)_t|(S_t)_t] = \prod_t \mathbb{P}[O_t|S_t]$$

Parameters (Explicit Duration)

- Emission $\mathbb{P}_\Theta[O_t|S_t]$
- Sojourn duration $\mathbb{P}_\Theta[d_i|X_i]$
- Transition $\mathbb{P}_\Theta[E_{i+1}|X_i]$

Issues

- Parameter estimation Θ
- Restoration of hidden sequence, e.g. Viterbi

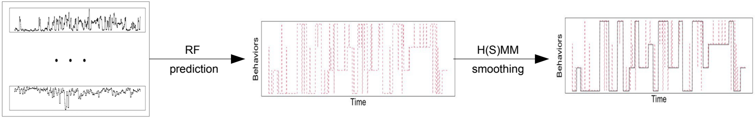
$$\arg\max_{(S_t)_{t\geq 1}} \mathbb{P}[(S_t)_{t\geq 1}|(O_t)_{t\geq 1},\Theta]$$

H(S)MM for accelerometer data

- In literature
  - HMM more frequent than HSMM

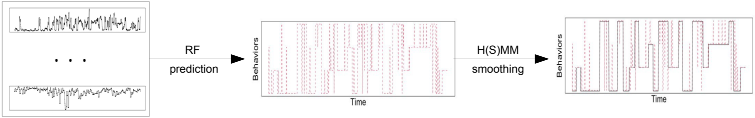
H(S)MM for accelerometer data

- In literature
  - HMM more frequent than HSMM
  - Observed variables : two strategies
    - ★ A/few selected feature(s) function of accelerometer (x,y,z)
    - ★ H(S)MM for smoothing RF predictions



H(S)MM for accelerometer data

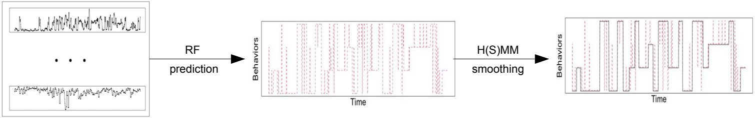
- In literature
  - HMM more frequent than HSMM
  - Observed variables : two strategies
    - A/few selected feature(s) function of accelerometer (x,y,z)
    - H(S)MM for smoothing RF predictions



  - Output = time budget/ accuracy /classification $\neq$ sequence analysis

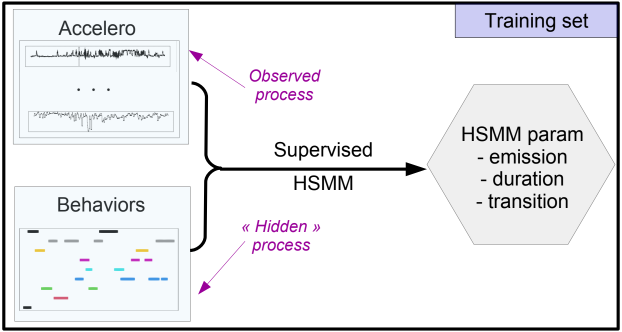
H(S)MM for accelerometer data

- In literature
  - HMM more frequent than HSMM
  - Observed variables : two strategies
    - A/few selected feature(s) function of accelerometer (x,y,z)
    - H(S)MM for smoothing RF predictions

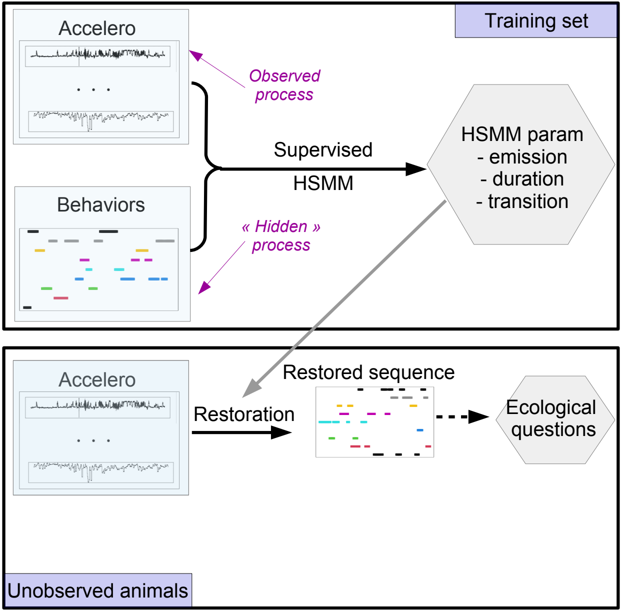


  - Output = time budget/ accuracy /classification $\neq$ sequence analysis
  - A few alternative works
    - Koslik *et al* (2025) : Inhomogeneous Markov models with interest on sojourn time distribution

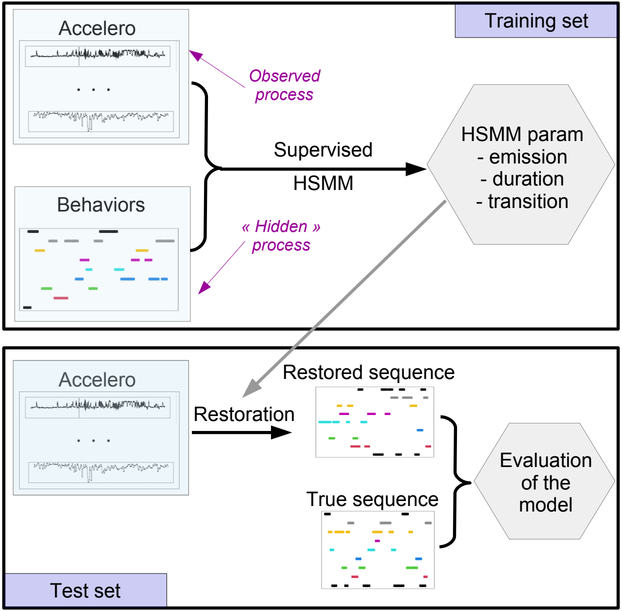
General statistical framework



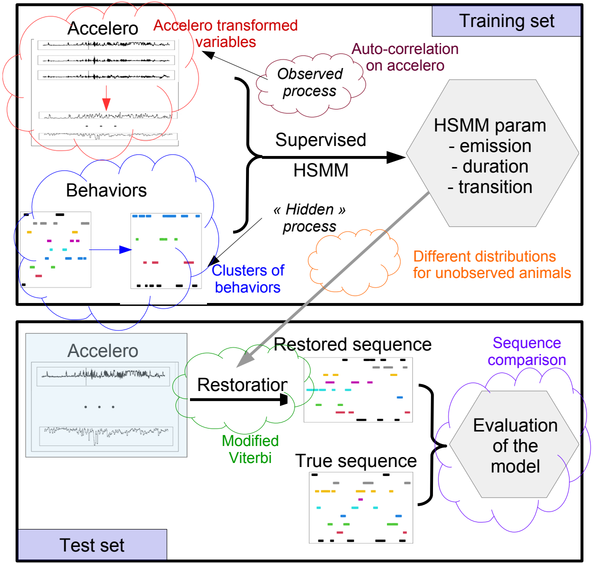
General statistical framework



General statistical framework

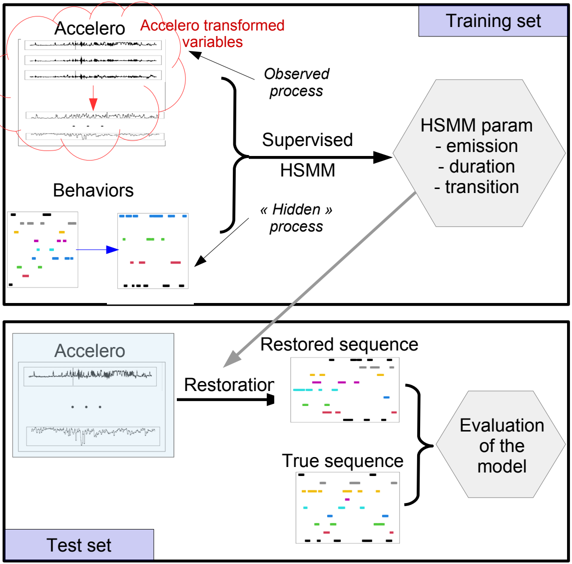


Various issues



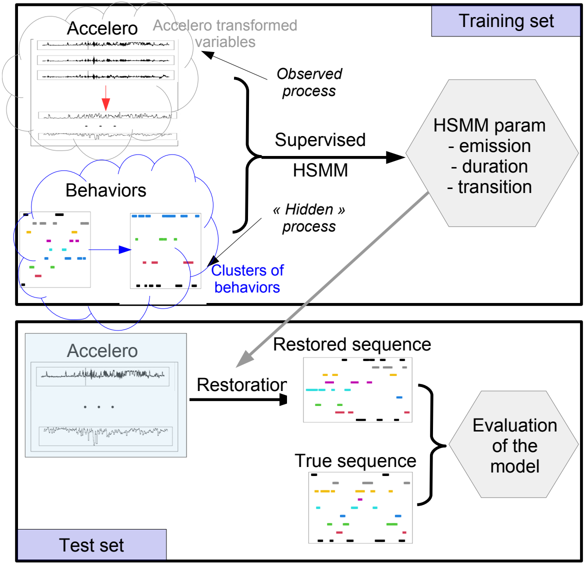
- Which accelerometer features
- Level of details of behaviors
- Autoregressive model for accelerometer data
- Modified Viterbi
- Metric for sequence similarity
- Robustness/transferability

Various issues



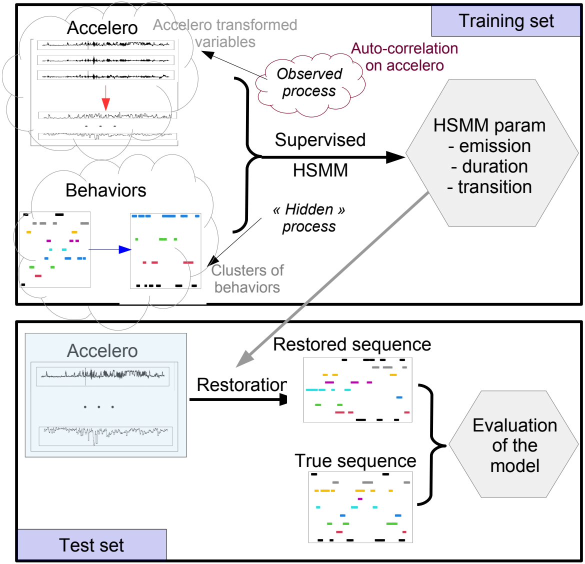
- Which accelerometer features
  - Balance complexity and discriminative power
- Level of details of behaviors
- Autoregressive model for accelerometer data
- Modified Viterbi
- Metric for sequence similarity
- Robustness/transferability

Various issues



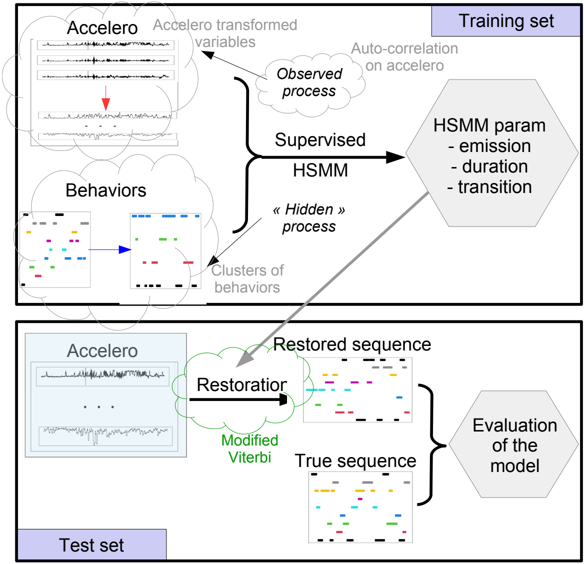
- Which accelerometer features
- Level of details of behaviors
  - distinguishable
  - interpretable
- Autoregressive model for accelerometer data
- Modified Viterbi
- Metric for sequence similarity
- Robustness/transferability

Various issues



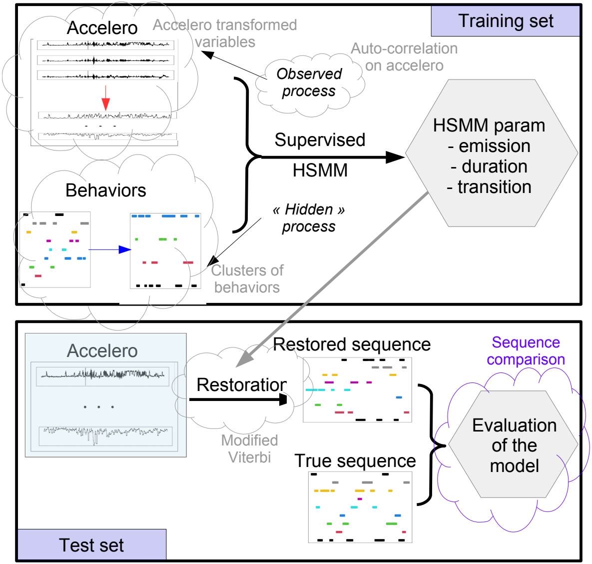
- Which accelerometer features
- Level of details of behaviors
- Autoregressive model for accelerometer data
  - Independent observations: unrealistic
  - HSMM-AR
- Modified Viterbi
- Metric for sequence similarity
- Robustness/transferability

Various issues



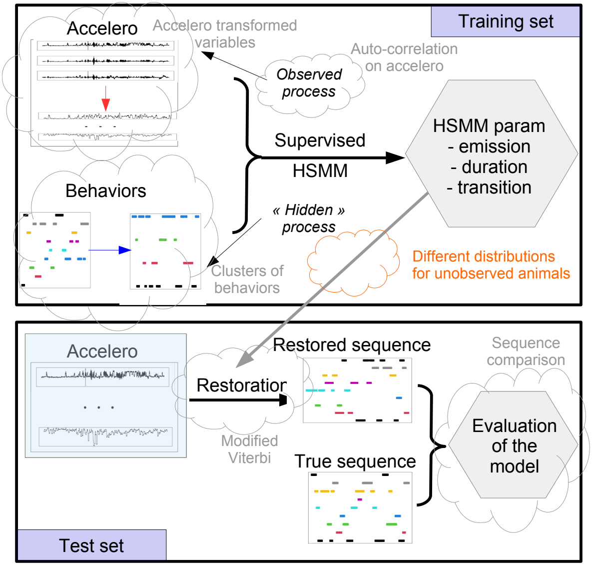
- Which accelerometer features?
- Level of details of behaviors
- Autoregressive model for accelerometer data
- Modified Viterbi
- Metric for sequence similarity
- Robustness/transferability

Various issues



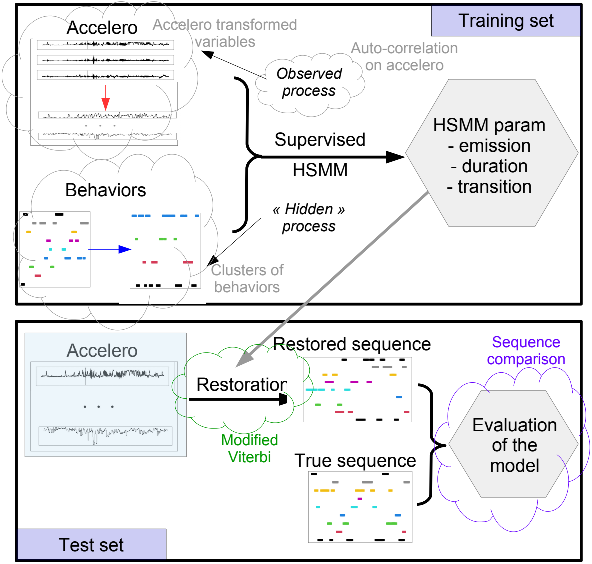
- Which accelerometer features
- Level of details of behaviors
- Autoregressive model for accelerometer data
- Modified Viterbi
- Metric for sequence similarity
  - What is a good restoration?
  - Also for hyperparameter calibration by CV
- Robustness/transferability

Various issues



- Which accelerometer features
- Level of details of behaviors
- Autoregressive model for accelerometer data
- Modified Viterbi
- Metric for sequence similarity
- Robustness/transferability
  - difference of behavior captive/wild animals

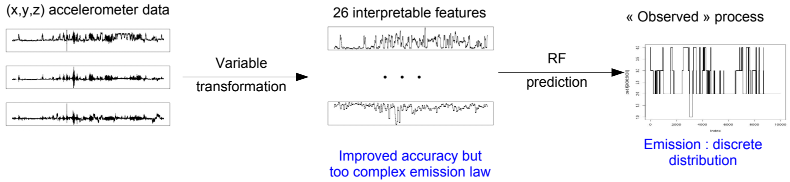
Various issues



- Which accelerometer features
- Level of details of behaviors
- Autoregressive model for accelerometer data
- Modified Viterbi
- Metric for sequence similarity
- Robustness/transferability

Model choices

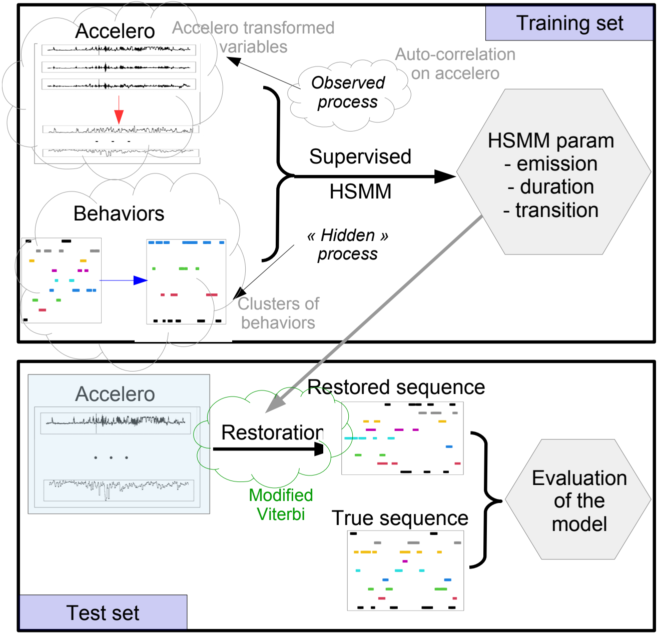
- "Observed" process = RF prediction



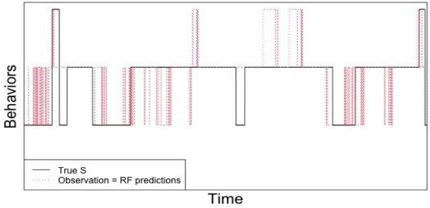
- 4 states $\leftrightarrow$ 4 clusters of behaviors
- Independent accelerometer observations
- Restoration by a modified version of Viterbi

Results

- Unsufficient quality of restoration for sequence-by-sequence analysis
  - Limits of the model but also of the information contained in the data
- Alternative goal : comparison between conditions (e.g. day/night)
  - Not done yet

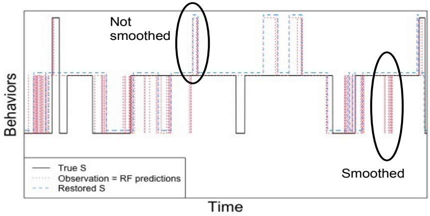

Training set
Accelero
Accelero transformed variables
Auto-correlation on accelero
Observed process
Supervised
HSMM
HSMM param
- emission
- duration
- transition
Behaviors
« Hidden » process
Clusters of behaviors
Accelero
Restored sequence
Restoration
Modified Viterbi
True sequence
Evaluation of the model
Test set

Restored sequences by Viterbi



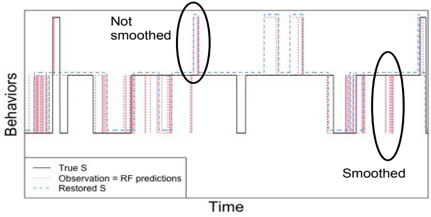
- RF : many wrong transitions and brief sojourns

Restored sequences by Viterbi

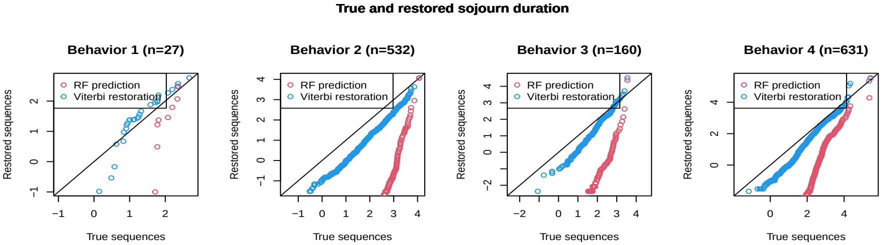


- RF : many wrong transitions and brief sojourns
- HSMM : partially smooth

Restored sequences by Viterbi



- RF : many wrong transitions and brief sojourns
- HSMM : partially smooth



- Over-representation of brief sojourn durations in restored sequences
- Brief sojourn durations unlikely given SMM parameters

  ⇒ Too much adequation to the observed process

Adequation to the model and to the observations

- Viterbi algorithm targets the most probable hidden sequence

$$\arg\max_S \log \mathbb{P}_\Theta[S|O] = \arg\max_S \{ \underbrace{\log \mathbb{P}_\Theta[O|S]}_{\substack{\textit{adequation}\\ \textit{to observations}}} + \underbrace{\log \mathbb{P}_\Theta[S]}_{\substack{\textit{adequation}\\ \textit{to model}}} + \cancel{\log \mathbb{P}_\Theta[O]} \}$$

Adequation to the model and to the observations

- Viterbi algorithm targets the most probable hidden sequence

$$\arg\max_S \log \mathbb{P}_\Theta[S|O] = \arg\max_S \{ \underbrace{\log \mathbb{P}_\Theta[O|S]}_{\substack{\textit{adequation}\\ \textit{to observations}}} + \underbrace{\log \mathbb{P}_\Theta[S]}_{\substack{\textit{adequation}\\ \textit{to model}}} + \cancel{\log \mathbb{P}_\Theta[O]} \}$$

- Modify the balance between adequation to the observations and to the model

$$\arg\max_S \{(1-\lambda) \log \mathbb{P}_\Theta[O|S] + \lambda \log \mathbb{P}_\Theta[S]\} \quad \text{with} \quad \lambda > 1$$

Adequation to the model and to the observations

- Viterbi algorithm targets the most probable hidden sequence

$$\arg\max_S \log \mathbb{P}_\Theta[S|O] = \arg\max_S \{ \underbrace{\log \mathbb{P}_\Theta[O|S]}_{\substack{adequation\\ to\ observations}} + \underbrace{\log \mathbb{P}_\Theta[S]}_{\substack{adequation\\ to\ model}} + \cancel{\log \mathbb{P}_\Theta[O]} \}$$

- Modify the balance between adequation to the observations and to the model

$$\arg\max_S \{(1-\lambda)\log \mathbb{P}_\Theta[O|S] + \lambda \log \mathbb{P}_\Theta[S]\} \quad \text{with} \quad \lambda > 1$$

- Select λ by cross-validation on the training set : which λ leads to the best restoration?
  - Metric for sequence comparison

(Digressive) remark: impact of sampling frequency on Viterbi

$$\arg\max_S \{\log \mathbb{P}[O|S] + \log \mathbb{P}[S]\}$$

Modify sampling frequency : $freq \to \lambda freq$

(Digressive) remark: impact of sampling frequency on Viterbi

$$\arg\max_S \{\log \mathbb{P}[O|S] + \log \mathbb{P}[S]\}$$

Modify sampling frequency : $freq \to \lambda freq$

- $\log \mathbb{P}[O|S] = \sum_{t=1}^{T} \log \mathbb{P}[O_t|S_t] \rightsquigarrow \lambda \log \mathbb{P}[O|S]$
  - time step $= 1/freq$

(Digressive) remark: impact of sampling frequency on Viterbi

$$\arg\max_S \{\log \mathbb{P}[O|S] + \log \mathbb{P}[S]\}$$

Modify sampling frequency : $freq \to \lambda freq$

- $\log \mathbb{P}[O|S] = \sum_{t=1}^{T} \log \mathbb{P}[O_t|S_t] \rightsquigarrow \lambda \log \mathbb{P}[O|S]$
  - time step $= 1/freq$
- $\log \mathbb{P}[S] = \sum_i \log \mathbb{P}_\Theta[(X_{i+1}, d_{i+1})|(X_i, d_i)] \rightsquigarrow \log \mathbb{P}[S]$
  - time step = successive states

Assumption : time step « sojourn duration

(Digressive) remark: impact of sampling frequency on Viterbi

$$\arg\max_S \{\log \mathbb{P}[O|S] + \log \mathbb{P}[S]\}$$

Modify sampling frequency : $freq \to \lambda freq$

- $\log \mathbb{P}[O|S] = \sum_{t=1}^{T} \log \mathbb{P}[O_t|S_t] \rightsquigarrow \lambda \log \mathbb{P}[O|S]$
  - time step = $1/freq$
- $\log \mathbb{P}[S] = \sum_i \log \mathbb{P}_\Theta[(X_{i+1}, d_{i+1})|(X_i, d_i)] \rightsquigarrow \log \mathbb{P}[S]$
  - time step = successive states

Assumption : time step « sojourn duration

$\Rightarrow$ The balance between the two terms is impacted

(Digressive) remark: impact of sampling frequency on Viterbi

$$\arg\max_S \{\log \mathbb{P}[O|S] + \log \mathbb{P}[S]\}$$

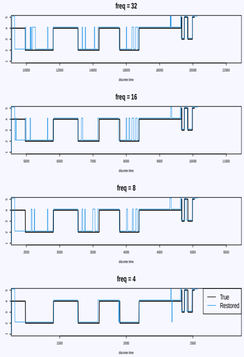
Modify sampling frequency : $freq \rightarrow \lambda freq$

- $\log \mathbb{P}[O|S] = \sum_{t=1}^{T} \log \mathbb{P}[O_t|S_t] \rightsquigarrow \lambda \log \mathbb{P}[O|S]$
  - time step $= 1/freq$
- $\log \mathbb{P}[S] = \sum_i \log \mathbb{P}_\Theta[(X_{i+1}, d_{i+1})|(X_i, d_i)] \rightsquigarrow \log \mathbb{P}[S]$
  - time step = successive states

Assumption : time step « sojourn duration

$\Rightarrow$ The balance between the two terms is impacted

Undersampling accelerometer data



(Digressive) remark: impact of sampling frequency on Viterbi

$$\arg\max_S \{\log \mathbb{P}[O|S] + \log \mathbb{P}[S]\}$$

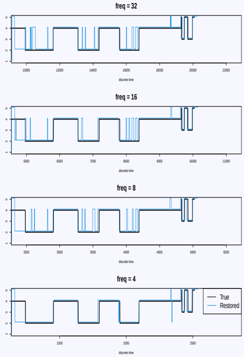
Modify sampling frequency : $freq \to \lambda freq$

- $\log \mathbb{P}[O|S] = \sum_{t=1}^{T} \log \mathbb{P}[O_t|S_t] \rightsquigarrow \lambda \log \mathbb{P}[O|S]$
  - time step $= 1/freq$
- $\log \mathbb{P}[S] = \sum_i \log \mathbb{P}_\Theta[(X_{i+1}, d_{i+1})|(X_i, d_i)] \rightsquigarrow \log \mathbb{P}[S]$
  - time step = successive states

Assumption : time step « sojourn duration

$\Rightarrow$ The balance between the two terms is impacted

Undersampling accelerometer data



$\Rightarrow$ Modified Viterbi is not only an artefact, it may also compensate the arbitrary choice of sampling frequency

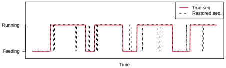
Similarity metric between sequences

$$\begin{cases} S = (S_t)_{t=1:T} & = \quad \text{true behaviour sequence} \\ \hat{S} = (\hat{S}_t)_{t=1:T} & = \quad \text{restored behaviour sequence} \end{cases}$$

Similarity metric between sequences

$$\begin{cases} S = (S_t)_{t=1:T} & = & \text{true behaviour sequence} \\ \hat{S} = (\hat{S}_t)_{t=1:T} & = & \text{restored behaviour sequence} \end{cases}$$

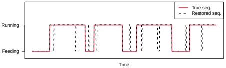
- L^p norms (euclidean,...) : $\left(\sum_t (S_t - \hat{S}_t)^p\right)^{1/p}$
  - Not adapted for analysis of durations/transitions



Similarity metric between sequences

$$\begin{cases} S = (S_t)_{t=1:T} & = & \text{true behaviour sequence} \\ \hat{S} = (\hat{S}_t)_{t=1:T} & = & \text{restored behaviour sequence} \end{cases}$$

- L^p norms (euclidean,...) : $\left(\sum_t (S_t - \hat{S}_t)^p\right)^{1/p}$
  - Not adapted for analysis of durations/transitions

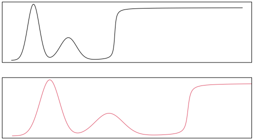


- Time series similarity measure requires
  - Define acceptable transformations. Example

Insertion/deletion (genomics)

ACTGGCAATCGAGTA
ACTGAGCAATCGAGT

Dilatation (finance)

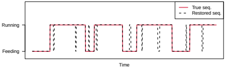


  - Tractable algorithm

Similarity metric between sequences

$$\begin{cases} S = (S_t)_{t=1:T} & = & \text{true behaviour sequence} \\ \hat{S} = (\hat{S}_t)_{t=1:T} & = & \text{restored behaviour sequence} \end{cases}$$

- L^p norms (euclidean,...) : $\left(\sum_t (S_t - \hat{S}_t)^p\right)^{1/p}$
  - Not adapted for analysis of durations/transitions

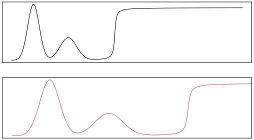


- Time series similarity measure requires
  - Define acceptable transformations. Example

Insertion/deletion (genomics)

ACTGGCAATCGAGTA
ACTGAGCAATCGAGT

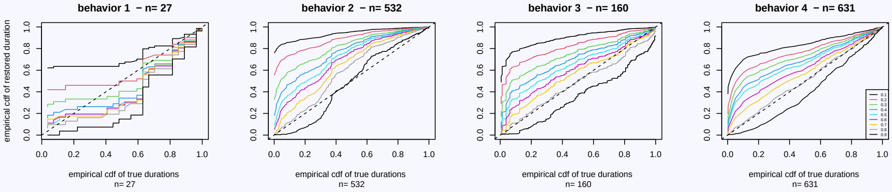
Dilatation (finance)



  - Tractable algorithm
- Similarity adapted to behavior sequence analysis?
  - Up to now : no lead for a sequence-by-sequence comparison
  - Compare duration and transitions **distributions**.

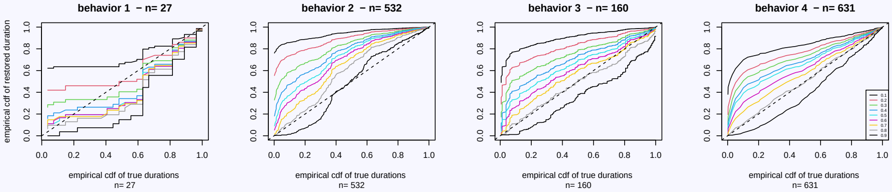
Choice of λ in modified Viterbi

- Average distance between empirical cdf of sojourn duration
- weight : number of occurence of this behavior in true sequences



Choice of λ in modified Viterbi

- Average distance between empirical cdf of sojourn duration
- weight : number of occurence of this behavior in true sequences

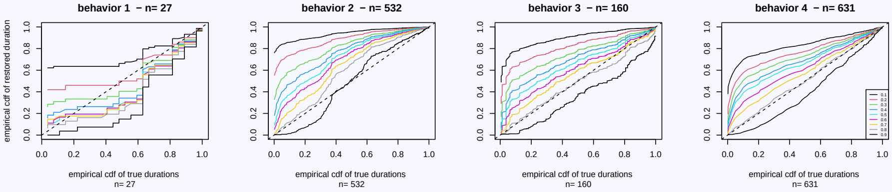


Limits

- Best λ may depends on the behavior class.
- Do not account for transition

Choice of λ in modified Viterbi

- Average distance between empirical cdf of sojourn duration
- weight : number of occurence of this behavior in true sequences



Limits

- Best λ may depends on the behavior class.
- Do not account for transition

Leads

- Metric adapted to a given ecological question (e.g. duration of a specific behavior) ?
- More generally : issue to be discussed with ecologists

Conclusion

- Goal : adresse new questions from accelerometer data w.r.t. classic analyses

Conclusion

- Goal : adresse new questions from accelerometer data w.r.t. classic analyses
- HSMM framework is not completely new but different outputs of interest
  - Impact various aspects of the model

Conclusion

- Goal : adresse new questions from accelerometer data w.r.t. classic analyses
- HSMM framework is not completely new but different outputs of interest
  - Impact various aspects of the model
- Work in progress: several axes of developments
  - First choice of modeling and parametrisation
  - Hope for better performances with model improvement

    $\hookrightarrow$ First axis : auto-correlation in accelerometer data

Conclusion

- Goal : adresse new questions from accelerometer data w.r.t. classic analyses
- HSMM framework is not completely new but different outputs of interest
  - Impact various aspects of the model
- Work in progress: several axes of developments
  - First choice of modeling and parametrisation
  - Hope for better performances with model improvement

    $\hookrightarrow$ First axis : auto-correlation in accelerometer data
- Fundamental limit: information brought by the data

  $\hookrightarrow$ Similar neck movements, behavior durations and transitions: impossible to distinguish.

Conclusion

- Goal : adresse new questions from accelerometer data w.r.t. classic analyses
- HSMM framework is not completely new but different outputs of interest
  - Impact various aspects of the model
- Work in progress: several axes of developments
  - First choice of modeling and parametrisation
  - Hope for better performances with model improvement

    $\hookrightarrow$ First axis : auto-correlation in accelerometer data
- Fundamental limit: information brought by the data

  $\hookrightarrow$ Similar neck movements, behavior durations and transitions: impossible to distinguish.
- Some issues at the interface of ecology and statistics