Proper account of long-term correlations in the observations improves state-space models' performances

Nicolas Bez, Pierre Gloaguen, Marie-Pierre Etienne, Rocio Joo, Sophie Lanco, Etienne Rivot, Emily Walker, Mathieu Woillez, Stéphanie Mahévas

,

BB-CY 4.0

Proper account of long-term correlations in the observations improves state-space models' performances

Nicolas Bez[1], Pierre Gloaguen[2], Marie-Pierre Etienne[3], Rocio Joo[4], Sophie Lanco[1], Etienne Rivot[5], Emily Walker[6], Mathieu Woillez[7], & Stéphanie Mahévas[1]

[1] Marbec, IRD, Ifremer, Univ Montpellier, CNRS – Sète, France
[2] Université de Bretagne Sud – Vannes, France
[3] IRMAR, Rennes University – Rennes, France
[4] Global Fishing Watch – Washington, USA
[5] Decod (Ecosystem Dynamics and Sustainability), L'Institut Agro, Ifremer, INRAE – Rennes, France
[6] BioSP, INRAE – Avignon, France
[7] Decod (Ecosystem Dynamics and Sustainability), L'Institut Agro, Ifremer, INRAE – Brest, France

Correspondence: nicolas.bez@ird.fr

Abstract

State-space models are widely used in ecology to infer hidden behaviors. This study develops an extensive numerical simulation-estimation experiment to evaluate the state decoding accuracy of four simple state-space models. These models are obtained by combining different Markovian specifications (Markov and semi-Markov) for the hidden layer with the absence and presence of auto-correlation for the observation layer. Model parameters are issued from two sets of real annotated trajectories. Three metrics are developed to help interpret model performance. The first is the Hellinger distance between Markov and semi-Markov sojourn time probability distributions. The second is sensitive to the overlap between the probability density functions of state-dependent variables (e.g., speed variables). The third quantifies the deterioration of the inference conditions between AR0 and AR1 formulations. It emerges that the most sensitive model choice concerns the auto-correlation of the random processes describing the state-dependent variables. Opting for the absence of auto-correlation in the model while the state-dependent variables are actually auto-correlated, is detrimental to state decoding performance. Regarding the hidden layer, imposing a Markov structure while the state process is semi-Markov does not impair the state decoding performances. The real-life estimates are consistent with our experimental finding that performance deteriorates when there are significant temporal correlations that are not accounted for in the model. In light of these findings, we recommend that researchers carefully consider the structure of the statistical model they suggest and confirm its alignment with the process being modeled, especially when considering the auto-correlation of observed variables.

Keywords: Hidden Markov models; Hidden semi-Markov models; state decoding accuracy; simulation-estimation; auto-regressive process; Ground truth

Introduction

The analysis of tracking data reveals information about the behavior underlying the movement patterns. After establishing the principle of representing behavior with a few meaningful states, segmentation methods can be used to estimate the states (Edelhoff et al., 2016). The commonly used segmentation methods includes filtering or thresholding (Gerritsen and Lordan, 2011), mixed models (Owen-Smith et al., 2010), K-means (Schwager et al., 2007), and hidden Markov models (HMM) or hidden semi-Markov models (HSMM) (Jonsen et al., 2013; Joo et al., 2013; Morales et al., 2004). This paper focuses on the two latter that allow time dependency to be considered (however, it is generally limited to first order, whereas data can have long temporal memories), and that are hierarchical models. Typically, they are composed of two layers: one to model the dynamics of the unobserved state process (also called the hidden process) and one to model the observed variables along the track (see, Zucchini and MacDonald, 2009 or McClintock, Langrock, et al., 2020 for didactic descriptions). Each of these layers can be flexible in their specifications (Gloaguen et al., 2015; Langrock et al., 2012; McClintock, King, et al., 2012; Morales et al., 2004). For example, one could use a Markov or semi-Markov hidden process to model the sequence of hidden states, and/or a correlated or uncorrelated random walk for the observed state-dependent variables.

When the states are known by direct or indirect observations (supervised cases), it is possible to infer properly the structure and the parameters of a segmentation model on a labeled training set, and to use the calibrated model to predict the states for a new unlabeled dataset. Learning methods, e.g., random forests (Sur et al., 2017; Thums et al., 2008) can also be used in these situations. However, when it comes to the most common unsupervised frameworks, model specification flexibility comes at a cost. Histograms for state-dependent variables and sojourn times in different states cannot be access to. As a result, one cannot infer the probabilistic laws they follow. In such cases, the *type* of probability density functions (PDF) of the state-dependent variables and the *type* of probability mass functions (PMF) of the sojourn times in the various states are not empirically grounded. They cannot be formally verified (Auger-Méthé, Clair, et al., 2011; Avgar et al., 2013; Pohle et al., 2017). However, postulating that the hidden process is Markov (HMM), with no time dependence in the observed state-dependent variables (AR0), may have a negative impact on the state decoding accuracy if the data do not conform with these choices. This study aims to assess the implications of some standard model specification choices.

In the absence of ground truth (unsupervised situations), a model can be evaluated against competing models using penalized maximum likelihood criteria such as AIC or BIC (Auger-Méthé, Newman, et al., 2021; Joo et al., 2013; Langrock et al., 2012; Pohle et al., 2017). Based on likelihood, selection is made from the standpoint of the visible side of the model, i.e., the observation layer. The hidden layers are not explicitly considered, even though they are the estimation's target. Another challenge is that several hidden state sequences can result in the same likelihood, making it difficult to select models based on likelihood. Zucchini and MacDonald, 2009 suggest to compare pseudo-residuals and the conditional probabilities to observe one particular observation knowing all observations, except the current one, in a Normal-Normal scales. Similarly Auger-Méthé, Newman, et al., 2021 use probability integral transform (PIT). However, these approaches only evaluate model consistency regarding the observed state-dependent variables. They do not permit evaluating model performance in terms of state estimations, which is the primary goal of these models. Simulation-estimation experiments are widely used in unsupervised situations to understand and evaluate the strengths and weaknesses of models (e.g., for the HMM context, McClintock, 2021). They are especially suitable for evaluating state decoding accuracy. Simulation-estimation entails assuming a true model, i.e., a model type and its parameters, which are then used to simulate sequences of states as well as realizations of the dependent variables (simulation step). The simulated dependent variables are then used as in real life unsupervised situations to re-estimate the parameters of the model, the *real* one or another model, and the states (estimation step).

The use of state-space models in fisheries is motivated by the fact a significant proportion of vessels are required by regulation to be equipped with GPS systems that allow for the observation of their trajectory at regular time steps. However, we don't know when they're fishing. The use of state-space models proved to be efficient to estimate where and when a vessel is fishing knowing its trajectories, and thus to monitor fishing pressure on certain fish stocks (Bez et al., 2011; Vermard et al., 2010). This results however in uncontrolled estimation precision and underlines the need for an evaluation of the state decoding performances of such models. We were fortunate to have access to a few fishing vessel tracks where the fishing operations were meticulously documented. In this paper, we conduct simulation-estimation experiments with four types of models, the parameter values of which are derived from these sets of real trajectories, and compare the estimation of the states with the real ones. In addition to the fact that behavioral state durations must be longer than the observation time step, understanding an individual's behavior from his or her trajectory depends on the model's time step. This often coincides with the observation time step of the trajectory, which influences the temporal scope of correlations. Consequently, the ability of models to infer behavior is time-dependent, so it is useful to investigate the sensitivity of model performance to their temporal resolutions.

This study aims to assess state-decoding accuracy or the models' ability to recover true hidden states. Following a general introduction to the main characteristics of hidden Markov and semi-Markov models and to the three metrics defined in this study to compare them, we focus on four types of models by combining two Markovian structures for the hidden layer (Markov or semi-Markov) with two auto-regressive orders for the observation layer (AR0 or AR1). In line with the literature dealing with model skill assessment, i.e., on the ability to assess the fidelity of model behavior to truth (Allen and Somerfield, 2009; Lynch et al., 2009), we assess different facets of state decoding accuracy simulation-estimations experiments. We perform a Model Skill Assessment (MSA), where the model used for the simulation and the estimation is the same. Then, we carry out a Model Robustness Assessment (MRA) where the model used for estimation is one of all the possible models except the model used for simulation. Thanks to the ground-truther fishing behavior associated with the vessel trajectories, we finally complete the analyses by checking the model assumptions and, more importantly, evaluating the ability of the four types of models to recover the known states.

Material and methods

Definition of the simulation models

Hidden Markov (HMM) and hidden semi-Markov models (HSMM) are two similar state-space models formulated and implemented here in discrete regular-time frameworks. Table 1 lists all notations.

The hidden layer hosts a random process S_t taking N_S discrete possible values that can be considered as a Markov (resp. semi-Markov) process for HMM (resp. HSMM) state-space models. Given that we want to estimate one specific behavioral state rather than all of them, we consider two states ($N_S = 2$). This is also consistent with the data that are used to define the models' settings. In general, HMM cannot be considered a particular case of HSMM. As explained below, the only case in which an HMM can be considered a subset of an HSMM is when the probability mass functions (PMFs) for the HSMM's sojourn time are chosen from the negative binomial distributions.

The eventually multiple observation layers host the state-dependent variables. Two observed state-dependent variables are considered here ($N_V = 2$) corresponding to the average persistent and rotation speeds associated with a pair of consecutive locations and located by convention on the first of these points (Figure 1, Gloaguen et al., 2015). Formally, the properties of the average velocity are strongly dependent on the time delay between observations, so it is recommended in practice to take regular time steps. It also leads us to

investigate model performance as a function of time resolution.

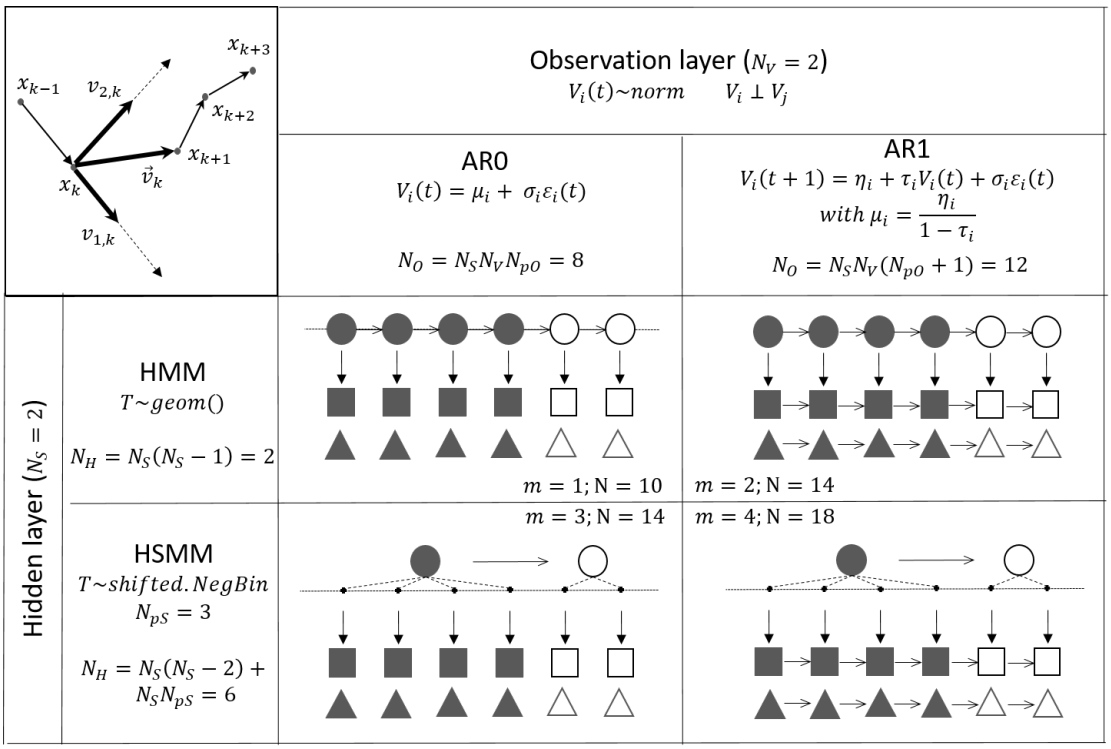

Figure 1. Directed acyclic graph for HMM and HSMM. The black arrows indicate conditional dependencies. Conditional independence is thus reflected by the absence of an arrow. The four tested models (m = 1,...,4) get one hidden layer with two states ($N_S = 2$) and one observation layer with two variables ($N_V = 2$) whose definition is indicated on the top-left panel. Circles represent the two states (white for State 1 and plain for State 2). The observation layer is made of two random variables depicted by squares and triangles. Conditionally on the states, the variables of the observation layer can follow an AR0 (no arrow between them) or an AR1 (arrows). The number of parameters needed by layer assumed a shifted-negative binomial PMF for the sojourn times in the HSMM formulation and a Gaussian PDF for the state-dependent variables. N_H and N_O represent the number of parameters required to parameterize the hidden and the observation layers, respectively.

Definition and characterization of the hidden layers

For HMMs, the Markov property (of order 1) implies that S_{t+1} is independent of $(S_1, S_2, \ldots, S_{t-1})$ conditionally on S_t, which means that:

$$[S_{t+1}|S_t = s_t, \ldots, S_1 = s_1] = [S_{t+1}|S_t = s_t]$$

where, $[X|Y = y]$ stands for the probability density function (PDF) of the random variable X, knowing that the random variable Y equals y. The stationary PMF of the sojourn times in the different states is defined by the transition matrix whose elements $p_{i,j}$ are the probabilities to switch from state i to state j when going from one time step to the next. A consequence of the conditional independence is that the sojourn time T_s in each possible state $s = 1, \ldots, N_S$ follows a geometric PMF:

$$T_s \sim geom(1 - p_{s,s})$$

where, $p_{s,s}$ is the probability of remaining in state s when going from one time step to the next. The first two moments are fully determined by $p_{s,s}$ as:

$$\mathbb{E}(T_s) = \frac{1}{(1 - p_{s,s})}$$
$$\mathbb{V}(T_s) = \frac{p_{s,s}}{(1 - p_{s,s})^2}$$

and the number of parameters required to specify the hidden layer of an HMM is $N_H = N_S(N_S - 1)$.

For HSMMs, the hidden layer itself can be decomposed into two sub-layers (R Core Team (2022), Figure 1). The first sub-layer corresponds to a Markov chain, not indexed on time, defined by a transition matrix that drives the realization of the states' sequence (the transition matrix gets a zero diagonal and lines summing to one so that it gets $N_S(N_S-2)$ elements). The second sub-layer specifies the PMF for the sojourn times in each state, not necessarily geometric and not necessarily parametric even though non-parametric formulations are rare (Johnson and Willsky, 2013; Sonia Malefaki and Limnios, 2010). In parametric cases, the hidden layer of an HSMM is defined by $N_H = N_S(N_S - 2) + N_S N_{pS}$ parameters, where N_{pS} represents the number of parameters defining the PMFs of the sojourn time. In this work, the PMFs for the sojourn time are chosen to be shifted negative binomial ($s\mathcal{NB}$) functions ($N_{pS} = 3$):

$$T_s \sim shift_s + \mathcal{NB}(n_s, p_s)$$

with

$$\mathbb{E}(T_s) = shift_s + n_s \frac{1 - p_s}{p_s}$$
$$\mathbb{V}(T_s) = n_s \frac{1 - p_s}{p_s^2}$$

The rationale for using such PMF is twofold. Firstly, in practice, sojourn times last at least one time step and cannot be null (shift=1). It is also common practice to obtain tracking data at a finer resolution than the duration of the states so that the smallest stay in a state is likely to be equivalent to a certain number of time steps (shift > 1 or >> 1). Secondly, geometrical PMFs, i.e. PMFs of the sojourn times in HMM cases, are particular cases of negative binomial PMFs ($n_s = 1$). Under this formulation, the HMM models are thus nested in the HSMM models (HMM $\subset$ HSMM$_{s\mathcal{NB}}$), which is not true in general for other HSMM specifications, i.e., for PMFs different from the family of the negative binomial functions. The distance between a shifted negative binomial PMF and its equivalent geometric PMF is evaluated by the Hellinger distances between the two PMFs (Hellinger, 1909). For each state $s = 1, ..., N_S$, the Hellinger distance d_{T_s} is given by:

$$d_{T_s} = \sqrt{1 - \sum_{t=0}^{\infty} \sqrt{geom_s(t) \cdot s\mathcal{NB}_s(t)}}$$

Thus, the overall measure of distance between an HMM and an HSMM$_{s\mathcal{NB}}$ is:

$$d_T = \sum_s \pi_s d_{T_s}$$

where $\pi_s = \mathbb{E}(T_s) / \sum_{s'} \mathbb{E}(T_{s'})$ is the probability of being in state s.

Definition and characterization of the observation layer

Probability density functions As in standard practice, we consider that the state-dependent random variables $V_i, i = 1, 2$, are mutually independent ($V_{1,t} \perp V_{2,t'}, \forall t, t'$) with PDFs defined conditionally to the hidden process. In the present work, the two state-dependent variables are chosen uni-variate Gaussian ($N_{p0} = 2$):

$$[V_{i,t} | S_t = s] \sim \mathcal{N}(\mu_{i,s}, \sigma_{i,s}) = \mathcal{N}_{i,s}$$

Discrimination between PDFs The estimation of the hidden states is *a priori* all the easier when the speed distributions per state have no or low overlaps. As an alternative to a direct measure of overlap or distance between PDFs, the discrepancy between the PDFs is quantified here with regard to the quality of the state prediction they potentially permit (the larger, the better). It corresponds to the state decoding accuracy achieved when ignoring the hidden layer, i.e. when considering a mixed model (Figure 2). Indeed, in this simple case, state estimation entails selecting the most likely state based on speed values. The state decoding accuracy when considering both variables, considered independent in the current case, is defined as:

$$d_V = \sum_s \pi_s \iint P(\hat{S} = s|S = s, V_1 = v_1, V_2 = v_2)\mathcal{N}_1(v_1)\mathcal{N}_2(v_2)dv_1dv_2$$
$$= \sum_s \pi_s \iint 1_{\mathcal{N}_{1,s}(v_1)\mathcal{N}_{2,s}(v_2) > \mathcal{N}_{1,s'}(v_1)\mathcal{N}_{2,s'}(v_2)}\mathcal{N}_{1,s}(v_1)\mathcal{N}_{2,s}(v_2)dv_1dv_2$$

Using the empirical speed frequency distribution for variable i in state s yields empirical equivalents. These metrics, ranging from 0 to 1, serve as summary statistics for the observation layers of the various models used in this study.

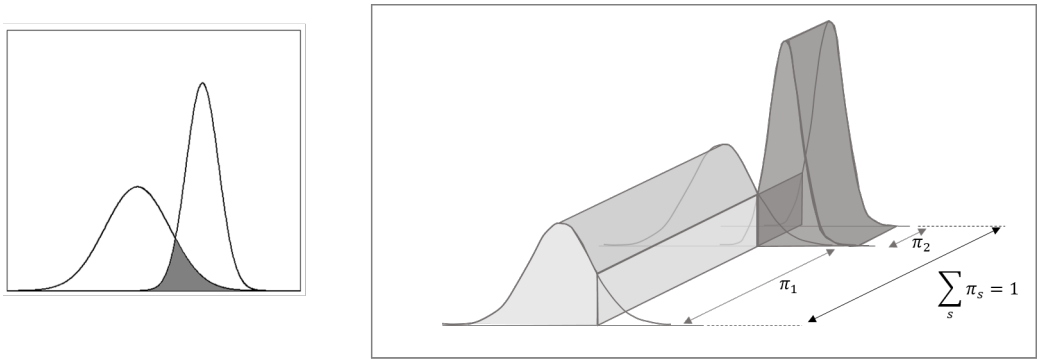

Figure 2. Overlap between PDFs and accuracy of state decoding assuming a simple mixed model with one single state-dependent variable. Each PDF is Gaussian. Knowing the variable value, the probability of getting the correct state corresponds to the PDF that is maximum weighted by the probability of being in the corresponding state.

Auto-regressive process of order 0 (AR0) It is also standard to consider that, conditionally to the state, $V_{i,t}$ and $V_{i,t+1}$ are not temporally correlated (auto-regressive process of order 0; AR0):

$$cor(V_{i,t}, V_{i,t+1}|S_t = S_{k+1} = s) = \tau_{i,s} = 0$$

The number of parameters characterizing the PDF of the state-dependent variables, the observation layer of an AR0, is defined by $N_O = N_S N_V N_{pO}$ parameters.

Auto-regressive processes of order 1 (AR1) A natural extension is to consider auto-regressive processes of order 1 (AR1) for $V_{i,t}$ (Gloaguen et al., 2015; McClintock, Langrock, et al., 2020; Poritz, 1982). Conditionally to a given state, say $S_t = S_{t+1} = s$, the auto-regressive processes can be formulated as follow: $V_{i,t+1} = \eta_{i,s} + \tau_{i,s}V_{i,t} + \sigma_{i,s}\epsilon$, where $\epsilon \sim \mathcal{N}(0, 1)$ so that

$$[V_{i,t}|S_t = s] \sim \mathcal{N}(\mu_{i,s} = \frac{\eta_{i,s}}{1 - \tau_{i,s}}, \sigma_{i,s})$$

For AR1, the number of parameters characterizing the PDF of the state-dependent variables is $N_O = N_S N_V (N_{pO} + 1)$. The extra parameter corresponds to the auto-regressive coefficient. AR0 being a particular case of AR1

(AR0 ⊂ AR1), it is possible to estimate an AR1 on the basis of AR0 data expecting that the coefficient of correlation will be 0 ($\tau_{i,s}$ = 0), while the reverse is unwise.

The discrete auto-correlation function of an AR1 writes $\rho(t) = \tau^t$. The correlation time (Yaglom, 1987) or the integral range (Lantuéjoul, 1991) is defined by $A = \sum_{t=0}^{\infty} \tau^t$. It is homogeneous to a time duration and characterizes the intrinsic temporal scale of a process ($A = 1$ unit of time for a nugget effect, $A = 2$ when $\tau = 0.5$ and $A = 10$ when $\tau = 0.9$). In terms of prediction, it contributes to define if a given observation time window can be regarded as large enough (resp. too low) for the mean of one realization to represent a precise prediction the mean of the process. This says that the time window of observation must be 10 times larger for an AR1 with $\tau = 0.9$ than for an AR0 to get the same level of precision when estimating the mean of the process. Reversely, when the time window is fixed, the precision deteriorates with increasing correlation coefficients.

The metric used to measure the difference between the AR1 and AR0 formulations is defined as the difference between the ratio of the integral range to the mean time duration of the various states. For an AR0 state-space model, the integral range is 1, and the ratio is simply $1/\mathbb{E}(T_s)$. The unit-less differences by variable and state ($d_{AR_{i,s}}$), by variable (d_{AR_s}) and for all variables and states (d_{AR}) between AR1 and AR0 formulations of a given state-space models are defined in this study by:

$$d_{AR_{i,s}} = \frac{A_{i,s} - 1}{\mathbb{E}(T_s)}$$

$$d_{AR_s} = \frac{1}{N_V} \sum_i d_{AR_{i,s}}$$

$$d_{AR} = \sum_s \pi_s d_{AR_s}$$

Definition of the simulation models

Four model types Four types of models are investigated (Figure 1) by crossing the two types of Markov structures with the two types of observation processes:

m_1 = HMM-AR0 $\qquad$ m_2 = HMM-AR1

m_3 = HSMM-AR0 $\qquad$ m_4 = HSMM-AR1

The number of parameters ($N_\theta = N_H + N_O$) ranges from 10 to 18 depending on the model.

Model parameters and time resolutions Parameters are set using ground-truthed tracks from two different fishing vessels with distinct fishing behaviors. Vessel 1 is a trawler operating in the North-East Atlantic with onboard sensors that detect fishing operations using GPS acquisition at 0.001 Hz (every 15 minutes). Vessel 2 is a purse seiner fishing in the Indian Ocean, with observers on board who record when the vessel is fishing and GPS acquisition at 0.003 Hz (every 5 minutes). For both vessels, the reported states are non-fishing (State 1) and fishing (State 2). Persistent and rotation speeds are calculated at four different time resolutions by degrading observations at 30 minutes, 1 hour, and 2 hours for Vessel 1 and 15 minutes, 30 minutes, and 1 hour for Vessel 2, respectively. To facilitate model comparison and gain generality, the resolution is expressed as the average number of time steps in the fishing state (i.e., State 2). For example, for Vessel 2, if a fishing operation lasts an hour and a half on average, the four resolutions correspond to 18, 6, 3, and 1.5 observations per average fishing operation.

For simulation-estimation experiments, parameter coherence across the four model types is critical. For the AR1 formulations, the parameters $\eta_{i,s}$ are deduced from $\tau_{i,s}$ and $\mu_{i,s}$ ensuring that the expected values and variances of the respective PDFs are the same in AR0 and AR1 for a given resolution. For the sojourn time PMFs, the parameters of the shifted negative binomial PMFs (*shift*,*n*,*p*) are evaluated by maximum likelihood based on observed tracking data. The HMM formulation's parameters are then deduced to ensure that the

mean durations are the same in both frameworks:

$$p_{s,s} = 1 - \frac{1}{shift_s + n_s \frac{1-p_s}{p_s}}$$

The variance is automatically deduced. The variances in state durations between the HSMM and HMM formulations may thus differ. Settings 1 (resp. 2) denote the sets of model parameters estimated from the tracks of Vessel 1 (resp. 2). Tables 2 and 3 list the model parameters.

Simulation-estimation experiments

Dimensions of the numerical experiments

Simulation-estimation experiments (Figure 3) assess the accuracy and robustness of state decoding under controlled conditions for:

2 settings × 4 resolutions × 4 model types = 32 different cases

For each model, 100 simulation-estimation experiments are run. In each experiment, synthetic data (states and speeds) are simulated, and the model parameters ($\hat{\theta}$) and states are re-estimated using expectation-maximization algorithms (refer to below). Each simulation lasts long enough to produce 250 state changes. We also carried out experiments with 50 alternations.

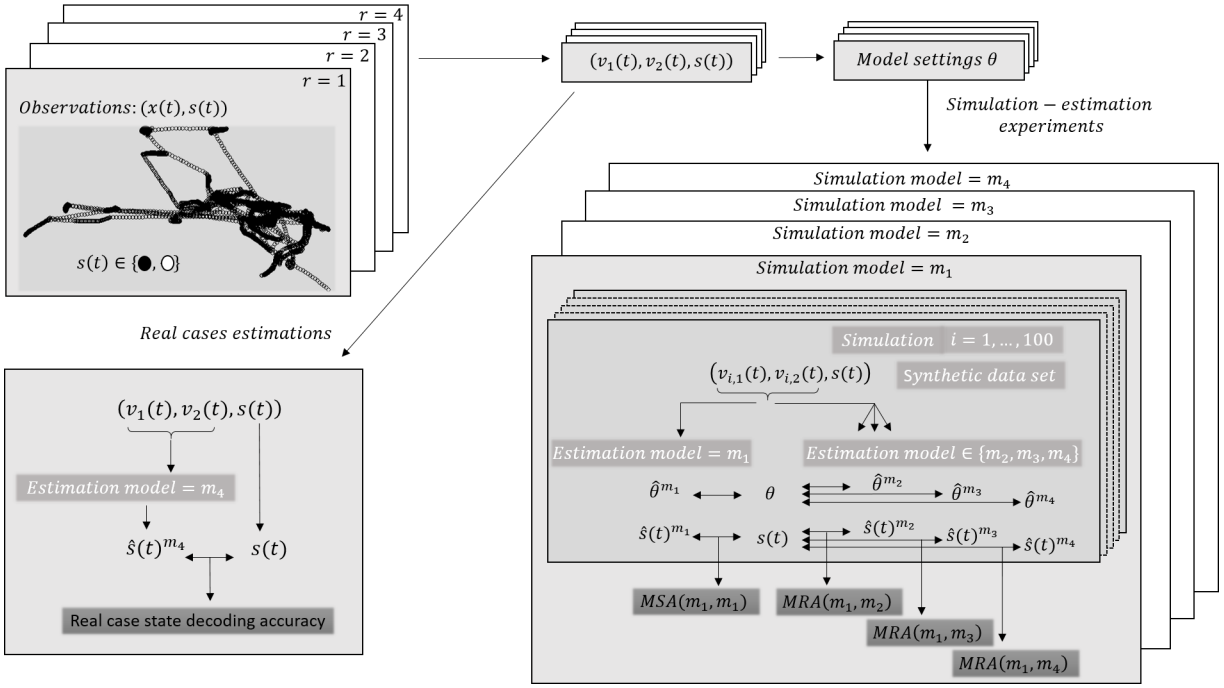

Figure 3. Workflow illustrated for one case study (grey panels represent the workflow for one particular resolution). ***Upper left panels***: observed time series in its original high-frequency resolution (r=1) and in three levels of up-scaling (r = 2,3,4). ***Upper right panels:*** computation of persistent and rotational speeds (v_1, v_2) for each resolution and estimation by maximum likelihood procedures of models' parameters and settings (θ). ***Bottom Right panels:*** the simulation-estimation experiments. For each tested model structure (m_1, m_2, m_3 and m_4), 100 data simulations are produced. For each simulation, the estimation is made using the EM algorithm with the same model (model skill assessments-MSA) or with a different model (model robustness assessment-MRA). Experiments get 250 or 50 alternations of states 1 and 2. ***Bottom left panel:*** estimation from the annotated real data with model structure m_4.

Maximum likelihood inference

Inferences are made by maximizing the likelihood using Expectation-Maximization (EM) algorithms (see Dempster et al., 1977 for HMM and Guédon, 2003 in the context of HSMM). EM algorithms are iterative and converge to the (possibly local) maximum likelihood. When the shape of the likelihood is unknown, it is recommended to run the algorithm from a variety of starting points. Inferences were made using starting values equal to ¼, ½, 1, 2, or 4 times the input parameter values used in simulations. Convergence occurred at the same values regardless of starting values, indicating that, in our case, the likelihood was unimodal, with a tractable optimum free of local and detrimental holes. To avoid mixing states, label switching (assigning State 1 to slow speed instead of large speed, and vice versa) is resolved prior to analyzing the results. The state estimate represents the maximum a posteriori at each time step. Alternatively, the Viterbi algorithm (Rabiner, 1989) can reconstruct the most likely sequence of states. In the current study, both outputs are very similar, so we only present the former.

State decoding accuracies

General definition The capacity to recover the model parameters (θ) is investigated using the box-plots of the hundred estimations of each of the model parameters. However, the performance is mainly based on state decoding accuracies as the objectives of the recourse to these models are to estimate states. Decoding accuracy is defined by the probability of correctly estimating the states (proportions of true positives and true negatives):

$$\text{accuracy} = P(\hat{S} = S) = \sum_s \pi_s P(\hat{S} = s | S = s)$$

Two facets of the accuracy are considered (Figure 3).

Model skill assessment (MSA) In sensitivity analysis (Allen and Somerfield, 2009), model skill assessment refers to scenarios where the simulation and estimation models are identical. This allows evaluating accuracy in the ideal conditions, that is when the data and model type are completely consistent (best case scenario):

$$\text{MSA}(m_i, m_i) = P(\hat{S} = S | \text{simulation model = } m_i, \text{estimation model = } m_i)$$

Model robustness assessment (MRA) In reference to the above MSA, the model robustness assessment (MRA) considers accuracies when the estimation model differs from the model used for simulation:

$$\text{MRA}(m_i, m_j) = P(\hat{S} = S | \text{simulation model = } m_i, \text{estimation model = } m_j)$$

For instance, if data are simulated according to m_4 = HSMM-AR1, estimations are performed with each of the three other possible model types, one by one (m_1, m_2, or m_3). An analysis of the impact of discrepancies between the Markov property of the hidden layer and/or the AR properties of the state-dependent layer allows testing which of the two discrepancies affects the most the percentage of loss of accuracy when using model m' instead of model m:

$$loss(m_i, m_j) = \frac{MRA(m_i, m_j) - MSA(m_i, m_i)}{MSA(m_i, m_i)}$$

Real case state decoding performance

HSMM-AR1, the most flexible model type, is fitted to two sets of real annotated trajectories (Figure 3) at different time resolutions. Real states are used to calculate the associated accuracy. The most likely order of auto-regressive processes used to model the speed time series per state is an important piece of information for interpreting real-world decoding performance. To test if a (partial) coefficient of correlation of order k is null, the appropriate Student test statistics is:

$$\frac{R}{\sqrt{\frac{1-R^2}{n-k-2}}} \sim T_{n-k-2}$$

where n represents the number of observations so that $n-k$ represents the number of pairs of data k apart in time.

In this study, the time series are interrupted by each change of state, and the lengths of each section are variables (Figure 4). So the number of data available to test the value of partial coefficients of correlation of order k is rather n_k, the number of pairs of observations available at time distance k accounting for the gaps in the time series. For each state-dependent variable and each state, the order of the auto-regressive process is estimated as the smallest order for which the empirical partial correlation falls within the 90% confidence intervals of an $T(n_k-2)$.

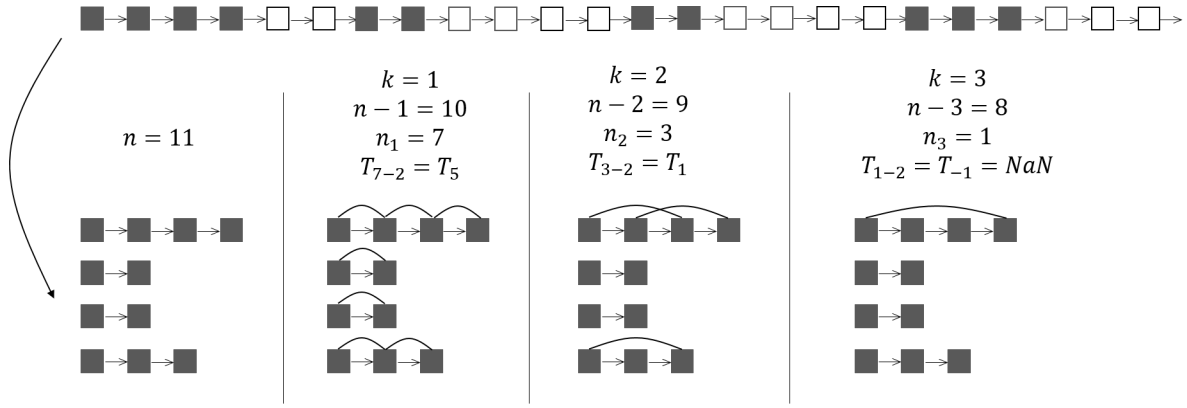

Figure 4. Adaptation of the Student T-test to account for the effective number of available data n_k for each tested order k for an auto-regressive process.

Results

Characteristics of the simulation models

Comparing settings

Figure 5 depicts the features that distinguish the two settings. In both settings, d_T and d_{AR} are co-linear, and they increase when the resolutions decrease (up to the exception of d_T for Settings 2, which is between the two highest resolutions). In other words, deviations from the HMM reference situation result in proportional deviations from the AR0 reference situation, which are significantly larger for Settings 1 than for Settings 2. Fitting an HMM-AR0 model to data simulated with an HSMM$_{s\mathcal{NB}}$-AR1 model is expected to be challenging, especially at high resolutions. The reverse is not a problem because AR0 and/or HMM models are special cases of AR1 and/or HSMM$_{s\mathcal{NB}}$ models. The proportions of time spent in states 1 and 2 are strongly unbalanced, with opposite dominance between settings. The state-specific characteristics of the various settings are thus weighted in the opposite direction. The average sojourn time in the various states also differs significantly between the two settings. In State 1, the average duration is ten times shorter in Settings 1 than in Settings 2 (regardless of resolution). When compared to State 2, it is 1.5 times larger. Despite the similarity of the correlation coefficients, the observed difference in d_{AR} can be attributed to this significant difference between the two settings. In State 1, the average correlation coefficients and integral ranges are very similar for the two settings (Figure 6, right column). We are using averages to explain the overall patterns; it is worth noting that the coefficients of correlation increase with increasing resolution. Despite similar coefficients of correlation, the distance d_{AR_1} for Settings 1 is 10 times larger due to their longer sojourn times. The picture for state two is different. Coefficients of correlation and integral ranges are slightly smaller in Settings 2. The slightly longer sojourn time results in a smaller d_{AR_2} for Settings 2. Therefore, d_{AR_1} and d_{AR_2} are larger for Settings 1. The unbalanced states' proportions between settings resulted in a larger d_{AR} for Settings 1 but reduced the

difference.

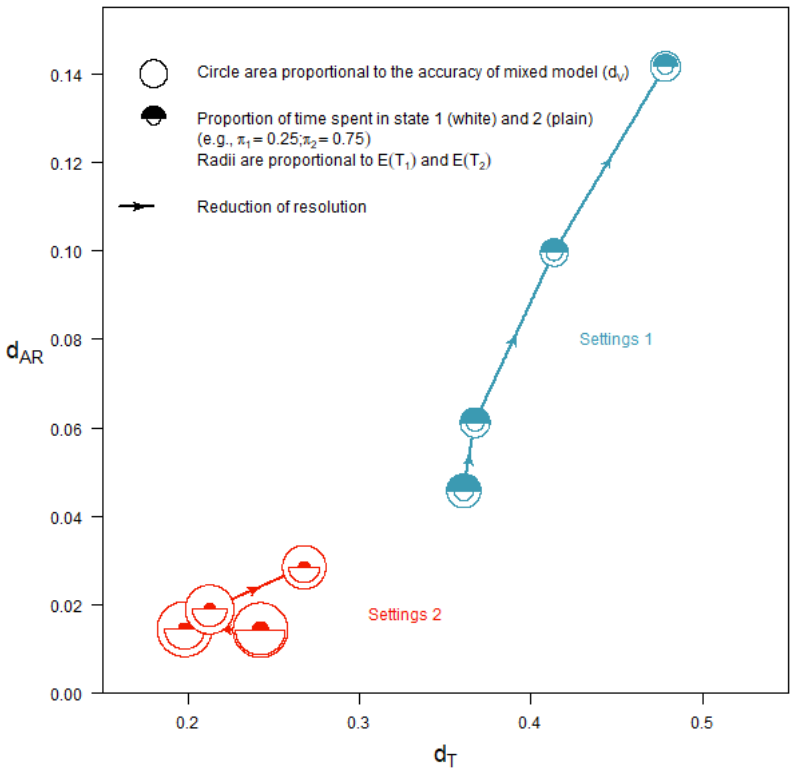

Figure 5. Graphical representation of the model settings. Settings 1 and 2 are summarized by d_T the Hellinger distances between the shifted-negative binomial (HSMM) and the geometric (HMM) PMF formulations for the sojourn time on the x-axis, by d_{AR} the difference between AR1 and AR0 formulations on the y-axis, and by d_V the state decoding accuracies in case of mixed model through the surface area of the circles. The proportions of time spent in each state are also represented by two half discs within the circle. The arrows in each setting indicate the path towards decreasing resolution.

Finally, given the overlaps between the PDFs of the speed variables by states, and the expected values of the different sojourn times, the expected state decoding capacities when using mixed models (d_V) are larger for Settings 2 than for Settings 1 (Figure 5), with 95% of good predictions for the largest resolution for Settings 2 as opposed to 80% for Settings 1 (Figure 9). All of these statements point to disappointing decoding accuracy for Settings 1 and decreasing resolutions in general. This latter statement suffers one exception: for Settings 1; the accuracy appears slightly better for the lowest resolution than the penultimate resolution (Figure 9). The expected values of the PMFs are equal in HMM and HSMM$_{s\mathcal{NB}}$ formulations, but the standard deviations differ. For Settings 1 and 2, the standard deviations are co-linear between the two types of Markov formulations but are generally larger in the HMMs (Figure 7). It also appears that, for Settings 1, the standard deviations of state durations in States 1 and 2 are of the same order of magnitude, whereas for Settings 2, they are one order of magnitude larger in State 1 than in State 2.

MSA (estimation model = simulation model)

Parameter estimation

In general, parameter estimates under MSA conditions are unbiased and symmetrically distributed around the target value (Figure 8). One notable exception is the variance of the Gaussian PDF for $V_{1,2}$ for Settings 2 when using an HSMM$_{s\mathcal{NB}}$-AR0 model. When the target value is 1.26, 50% of the estimations fall between $Q_{25\%} = 1.25$ and $Q_{75\%} = 1.44$, indicating a minor issue. The variance of the shifted negative binomial at

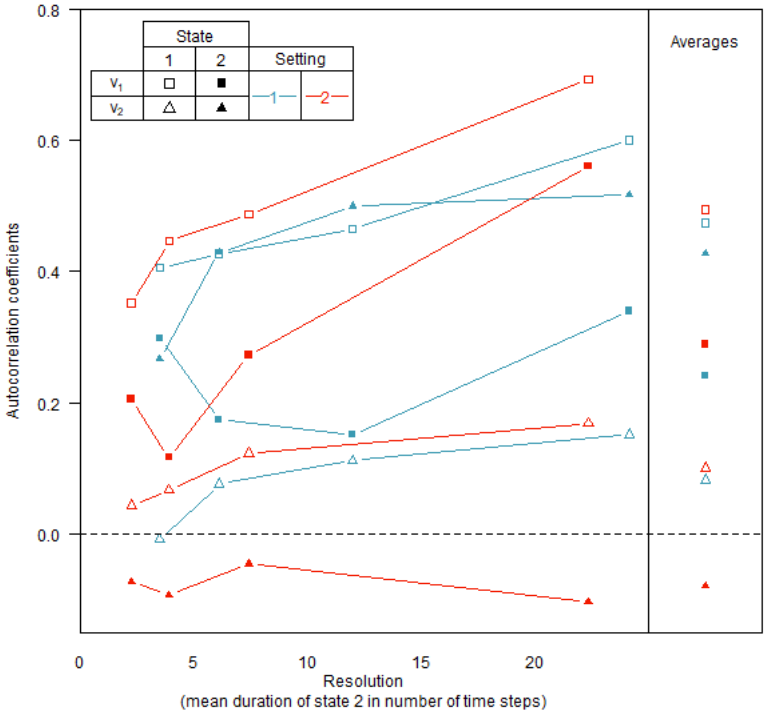

Figure 6. Auto-correlation coefficients for state-dependent variables in AR1 models, plotted against resolution. The average coefficients, regardless of resolution, appear in the right marginal column.

high resolution appears to be slightly positively biased as well (the mean of the estimation is slightly larger than the target value). Finally, for a given parameter, setting, and degradation, the ranges of fluctuations of the estimations change with the model used. This is to the disadvantage of HSMM$_{s\mathcal{NB}}$ frameworks (larger inter-quantile intervals).

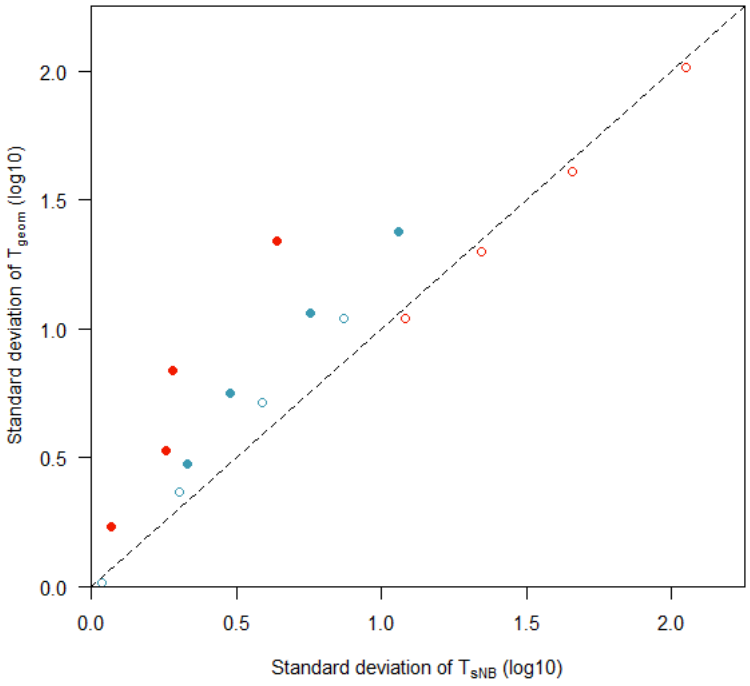

Figure 7. The decimal logarithm of standard deviations of the sojourn time PMFs for the various settings. The x-axis represents the time duration of HSMM cases with shifted negative binomial PMFs. The y-axis refers to HMM cases with geometric PMFs.

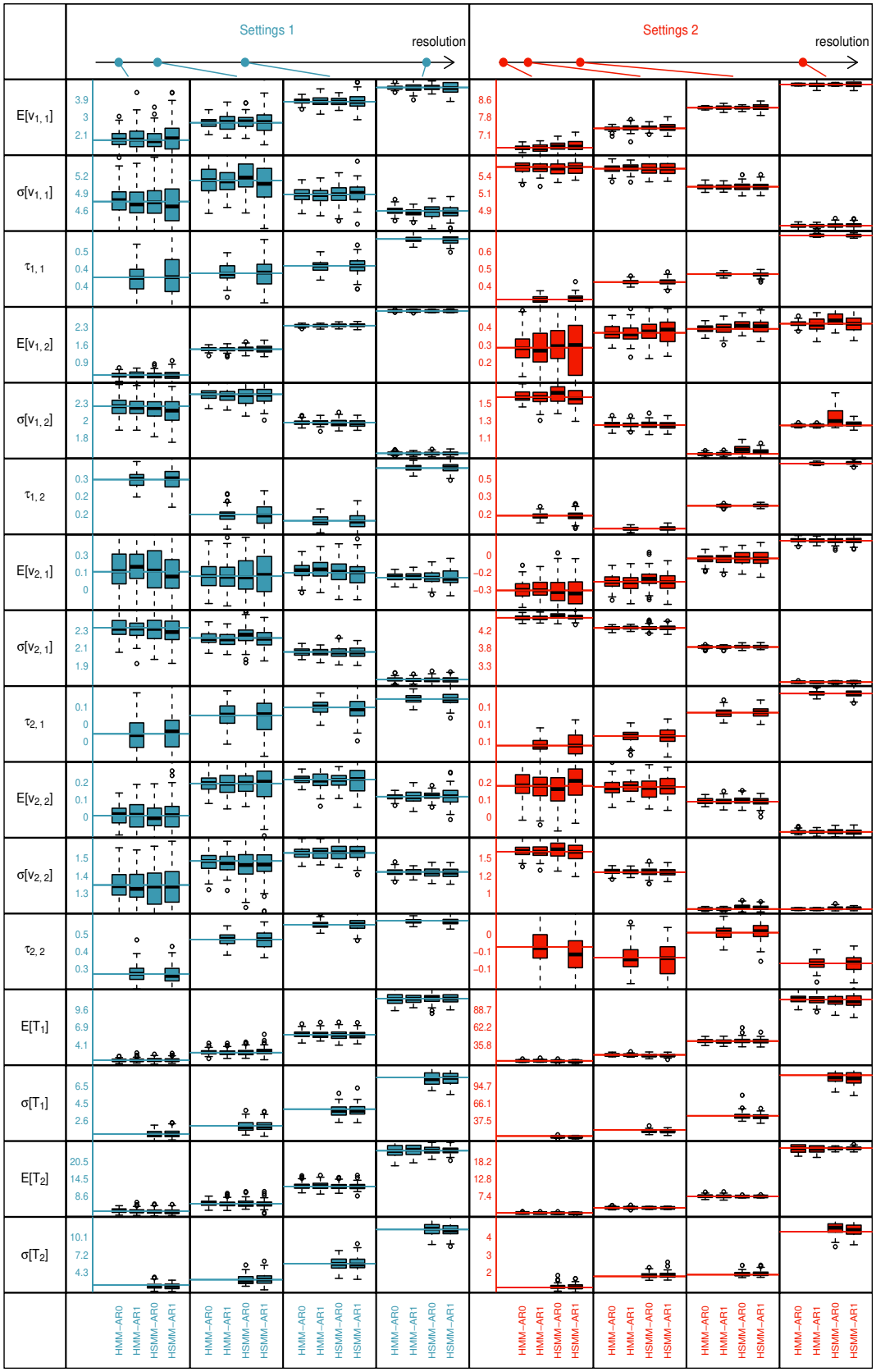

Figure 8. Simulation-estimation experiments: box-plots of the 100 parameter estimation when the model used for the estimation is the same as the model used for the simulation (MSA experiments with 250 states alternations). Each panel represents a settings × a degradation. The sequence of the box-plots is always the same: HMM-AR0, HMM-AR1, HSMM-AR0, HSMM-AR1. For each parameter, the y-axis is set to be consistent across all four resolutions for given settings. A horizontal line represents the simulation model parameter's (reference) value. The shift for the $s\mathcal{NB}$ PMF is not represented.

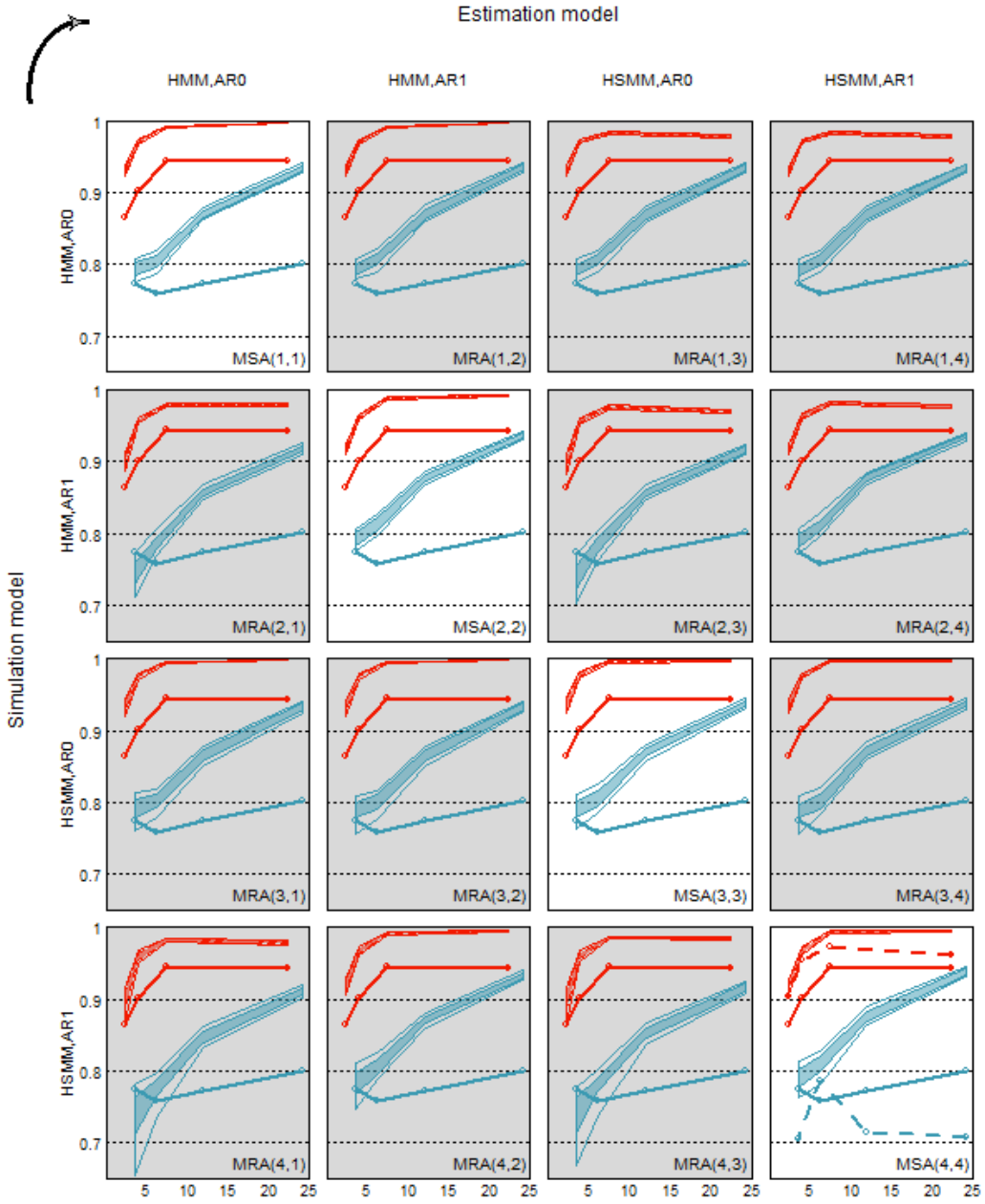

Figure 9. Accuracies (y-axis) of state estimations as a function of the resolution (x-axis expressed as the average number of time steps in State 2). ***Simulation-estimation experiments (with 250 states alternations)*** : in a row, the simulation models, and in the column, the estimation models. The light-colored envelopes are built on the basis of the 25% and 75% quantiles of the accuracies obtained over the 100 simulations performed for Settings 1 (blue) and 2 (red). The white envelopes correspond to the 10% and 90% quantiles. The white diagonal panels correspond to model skill assessment - MSA (estimation model equals the simulation model). The grey panels correspond to the model robustness assessment - MRA (estimation model equals all but the simulation model). ***Model characteristics***: the continuous lines indicate the expected accuracies when considering a mixed model (d_V). ***Real case accuracies:***: the dashed lines in the bottom right panel correspond to the accuracies for real data estimations when using an $\text{HSMM}_{s\mathbb{NB}}$-AR1.

State decoding accuracies

Figure 9 (diagonal panels) shows that state decoding accuracy is higher in Settings 2 compared to Settings 1, regardless of the model. In both cases, increasing the resolution (or the number of observations per state) improves accuracy. There is no difference between the model types. The state decoding accuracy is similar for the most complex models (HSMM-AR1) to the simplest ones (HMM-AR0), which outperform by around 10% what would be expected under a more simple mixed model framework. The AR1 process, when present in the data and model structure, serves as a source of improvement. It's worth noting that the results are consistent with those obtained using mixed models. For Settings 1, performance improves when moving from the smallest to the penultimate smallest resolution. Across the 100 simulations, the accuracy values are very similar. This is obvious for Settings 2 at all resolutions, demonstrating no difference between simulations. For Settings 1, some variations occur at low resolutions but remain small, with an average accuracy equal to 80%. These results serve as the baseline for state decoding performance because they are obtained under the most favorable conditions, with the estimation model equal to the simulation model. It is thus expected that performance will suffer in less favorable situations.

MRA (estimation model = all but the simulation model)

In MRA situations (non-diagonal panels of Figure 9 taken in a row), the overall picture remains the same: the accuracies are systematically larger for Settings 2 and get larger when the resolution increases. The largest changes in performances appear when the simulation model is AR1 and the estimation model is AR0 (Figure 9, panels of lines 2 and 4). The range of fluctuations of the accuracies remains similar to MSA conditions (very small for Settings 2 and small for Settings 1). As a result, average values in each numerical experiment provide a good summary of accuracy. The average relative accuracy losses (Figure 10) are shown below. The overall average loss is 1.2%. However, this average loss conceals contrasting situations, as follows.

Using the correct Markov structure (panels of the middle line), estimation with an AR1 has no impact (average loss equal to 0), which is consistent with the fact that an AR1 can capture an AR0 as a special case (AR0 ⊂ AR1). However, in the reverse situation (AR1 => AR0), accuracies deteriorate on average and significantly more for Settings 1. The loss of accuracy is linearly related to the distance between AR1 and AR0 formulations: the larger the distance d_{AR}, the greater the loss of accuracy (Figure 11).

Using the correct AR structure (panels of the middle column), an estimation with a HSMM$_{s\mathcal{NB}}$ when the data is generated with an HMM has little impact. Accuracies are, however, slightly tampered with for Settings 2, whose d_T values are nevertheless smaller than those for Settings 1 (Figure 5). These results support the idea that an HSMM$_{s\mathcal{NB}}$ can capture an HMM as a special case (HMM ⊂ HSMM$_{s\mathcal{NB}}$). The situation remains the same the other way round (HSMM$_{s\mathcal{NB}}$ => HMM) contrary to expectations. One particular case of Settings 1 is particularly harmful, while the rest have no effect.

Estimating with an AR1 structure when the data is AR0 does not affect the state decoding accuracy (left column). The most significant impacts occur when estimating using an AR0 structure while the true process is an AR1 (right column panels), and the value case is when also forcing the estimation to be HMM while the data are HSMM$_{s\mathcal{NB}}$ (average loss equals 3.6%). Changing from HSMM$_{s\mathcal{NB}}$ to HMM is only a concern when there is a negative change in the AR structure of the estimating model (i.e. AR0 instead of AR1). In this worst case, a fine inspection of the parameters' re-estimations when simulating with an HSMM-AR1 and estimating with one of the three other models (Figure 12) shows that nearly all parameters are impacted by this bi-layers discrepancy between the simulating and estimating model. What is also clear is that the main impact occurs when the state-dependent variables are modeled using the AR0 structure. For example, when an HMM-AR0 or an HSMM-AR0 is used to estimate the variance parameters of the Gaussian PDFs of V_1, there is a significant

bias. Expected values of sojourn time are also significantly affected. However, the most significant effects are observed when an HMM formulation is used for the estimation model.

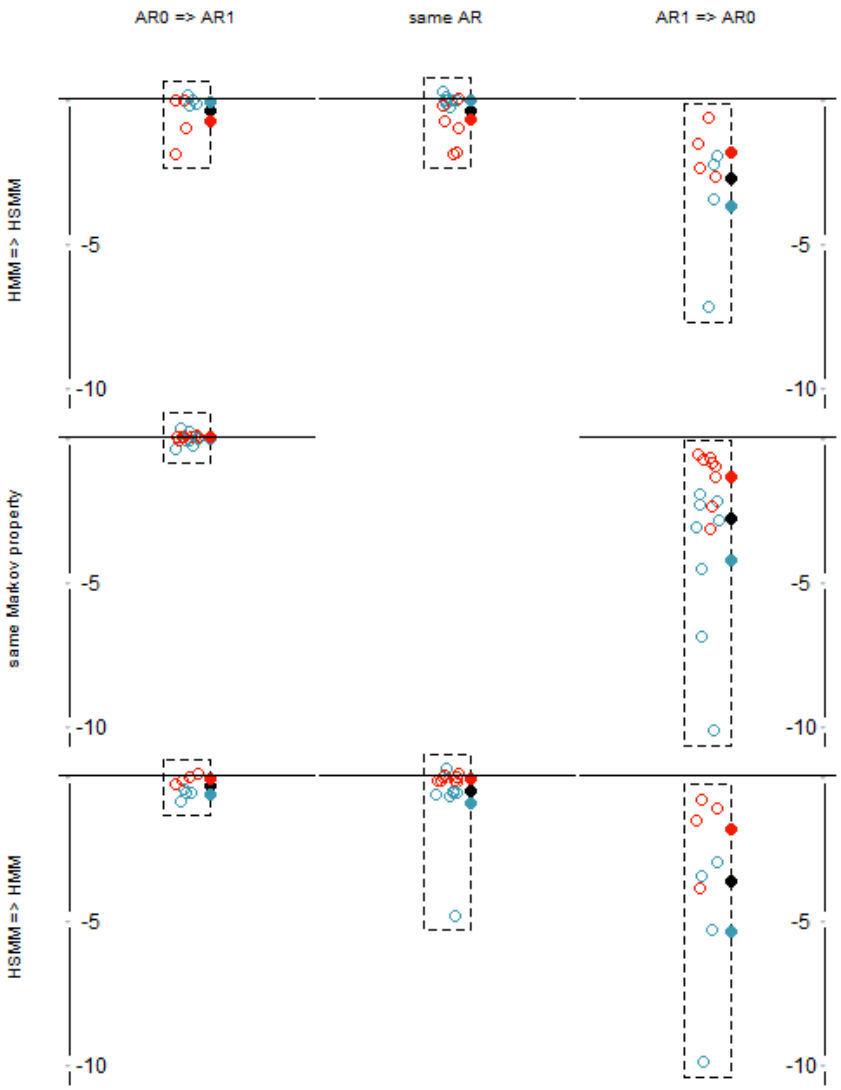

Figure 10. Model robustness assessment-MRA (with 250 states alternations). The percentage of state estimation accuracy lost when using a different estimation model than the one used to simulate the data. The boxes show the ranges of fluctuations. Points are jittered for better readability. The black dots on the right of each panel represent the overall mean of each case (blue and red for Settings 1 and 2, respectively). The changes in model structures between simulation and estimation are shown in the left and top margins. Columns represent a possible change in the auto-regressive characteristics of the speed distributions, e.g., "AR0 => AR1" means simulating with AR0 and re-estimating with AR1. Rows correspond to a possible change of the Markov property of the model, e.g., "HMM => HSMM" means: simulating with HMM and re-estimating with HSMM$_{s\mathcal{NB}}$. The central panel corresponding to the MSA situation is not considered.

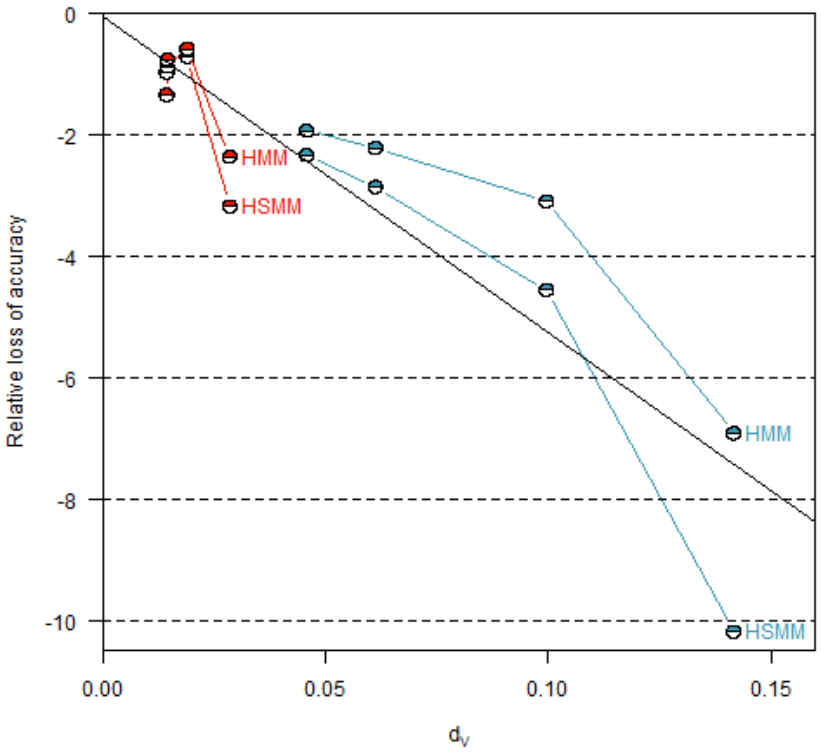

Figure 11. Simulation-estimation analyses-MRA (experiments with 250 states alternations). Loss of accuracy of state estimations when simulating with an AR1 model vs. estimating with an AR0 model as a function of the distance between AR1 and AR0 formulations for Settings 1 (blue) and 2 (red).

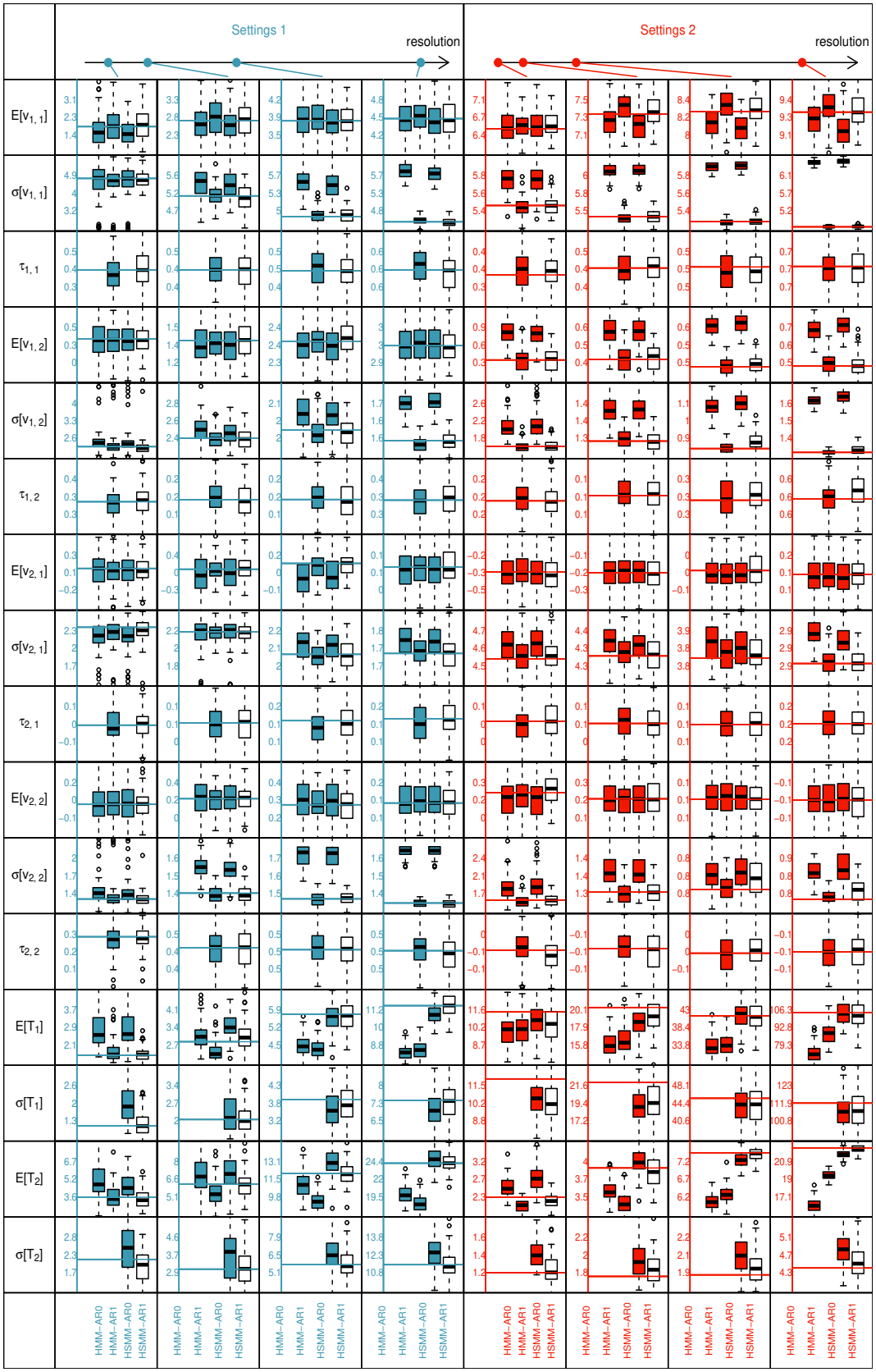

Figure 12. Simulation-estimation experiments: box-plots of the 100 parameter estimation when the model used for the estimation all but the HSMM-AR1 model used for the simulation (MRA experiments with 250 states alternations). Each panel represents a settings × a degradation. The sequence of the box-plots is always the same: HMM-AR0, HMM-AR1, HSMM-AR0, HSMM-AR1. The (reference) value of the parameter of the simulation model is represented by a horizontal line. The shift for the $s\mathcal{NB}$ PMF is not represented.

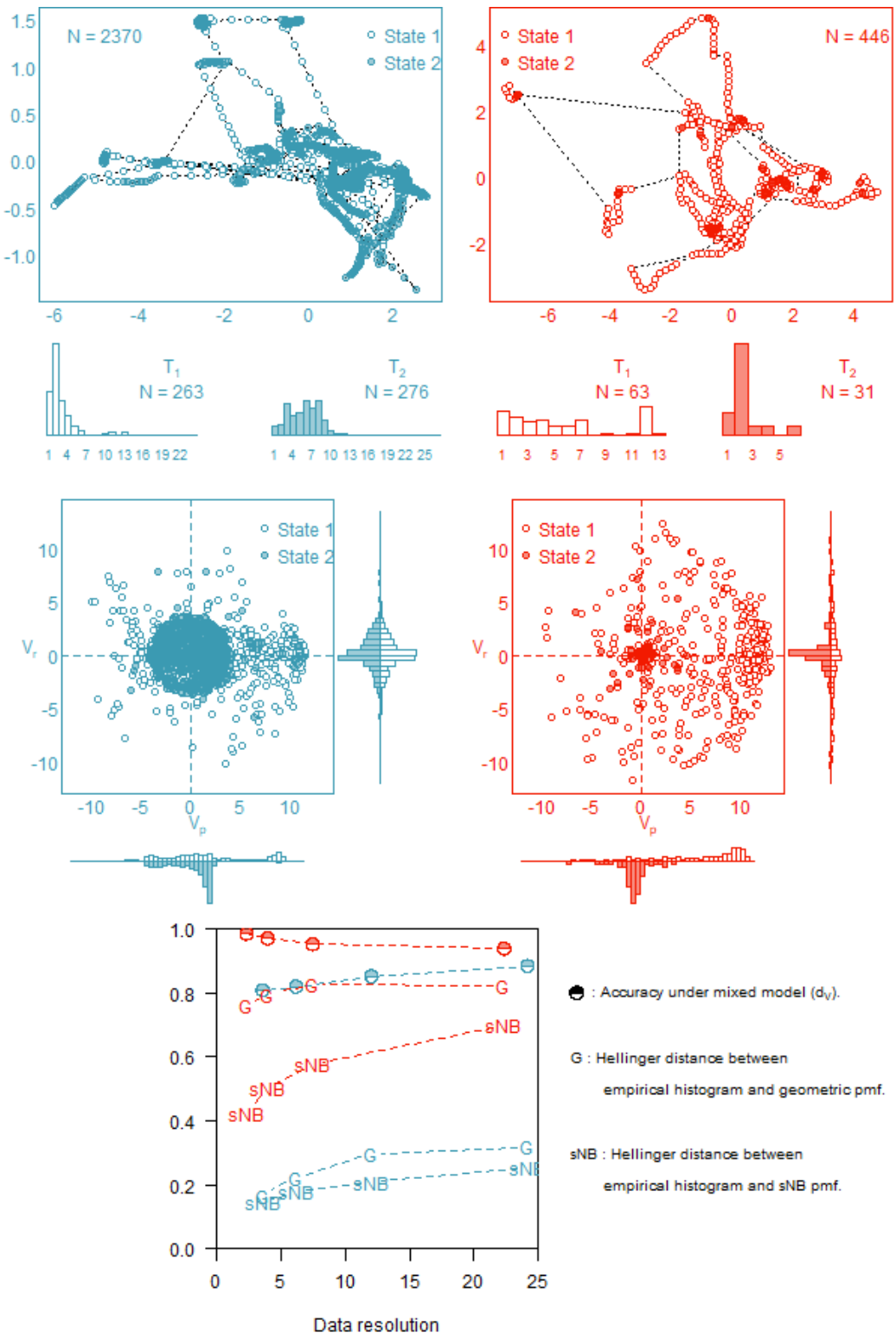

Figure 13. Data characteristics for Vessel 1 (blue) and Vessel 2 (red). State 1 (non-fishing) is in open symbols, and State 2 (fishing) is in plain symbols.

Top panels: real trajectories with centered coordinates for data resolution of 1 h. The histograms of the sojourn time are provided below.

Middle panels: correlation plots of the persistent (V_p) and the rotation (V_r) speeds. The marginal histograms represent the uni-variate distribution per state.

Bottom panel: summary metrics for the two vessels as a function of the data resolutions. The Hellinger distances between the sojourn time histograms and a geometric PMF ("G") or a shifted negative binomial PMF ("sNB"), and the accuracies expected under mixed model (d_V).

Real cases

Choosing a model structure and describing its main characteristics

Figure 13 shows the general characteristics of two sets of real tracks. As the sequences of true behavioral states are known, it is possible to select the best model by evaluating parametric assumptions for both the hidden state variables and the state-dependent variables. However, given a large number of observations (from several hundred to a few thousand depending on resolution), any minor difference between model and data characteristics results in low p-values for null hypothesis testing, even at low-risk levels ("large N problem").

Hidden layer

For Vessel 1, the Hellinger distances between the histograms of sojourn time duration in each state and either of the two possible PMFs to model the state process (i.e. geometric and shifted negative binomial) indicate that the most appropriate model is always an HSMM$_{s\mathcal{NB}}$ (Figure 13). For Vessel 2, this could be less relevant because the distances are greater. But choosing a geometric PMF would even be worth it.

Observation layer

The state-dependent variables exhibit temporal auto-correlation at lag 1 (Figure 6). Among the four structures tested in this work, the most appropriate model structure is an HSMM$_{s\mathcal{NB}}$ - AR1, which is also the most flexible one with regards to inference objectives as it can converge down to an HMM - AR0 if necessary. However, in the framework of auto-regressive processes, which is an assumption itself, the orders of the AR processes of the speed variables are likely larger than 1 (Figure 14). This is very clear for Vessel 1, which has auto-regressive processes with an order greater than 1 in half of the cases (9/16), as opposed to 5/16 cases for Vessel 2, and has a larger partial auto-correlation than Vessel 2. The correlation plots show that the persistent and rotation speeds are not independent given the state (except for Vessel 2 in State 2 , Figure 13) which contradicts the model's assumptions. Finally, Vessel 1 has strong mixing (respectively small d_V) between the empirical histograms of the speed in different states, particularly for the rotation speed (Figure 13), indicating poor state decoding performance.

State decoding performances

HSMM$_{s\mathcal{NB}}$ - AR1 models and speed-based filters are applied to the speeds time series available for the two vessels at the four different time resolutions (Figures 3) to estimate the (known but removed) states. Given the number of fishing operations (corresponding to State 2 by convention here) observed along each trajectory (N = 276 for Vessel 1 and N = 31 for Vessel 2; Figure 13), real state decoding accuracy is compared to the simulation-estimation experiments based on 250 and 50 alternations for Settings 1 and 2 respectively. Real-case estimates (Figure 9) perform worse than simulation-estimation experiments. This is especially true for Vessel 1 at high resolution where the real-case state decoding performances clearly depart from the simulation-estimation experiment (70% of good state estimation for the HSMM$_{s\mathcal{NB}}$ - AR1 model, compared to 90% for the simulation-estimation experiment). Although not shown in this paper, these models surpass the conventional filtering approach based on a speed threshold to routinely identify behavior for fisheries management.

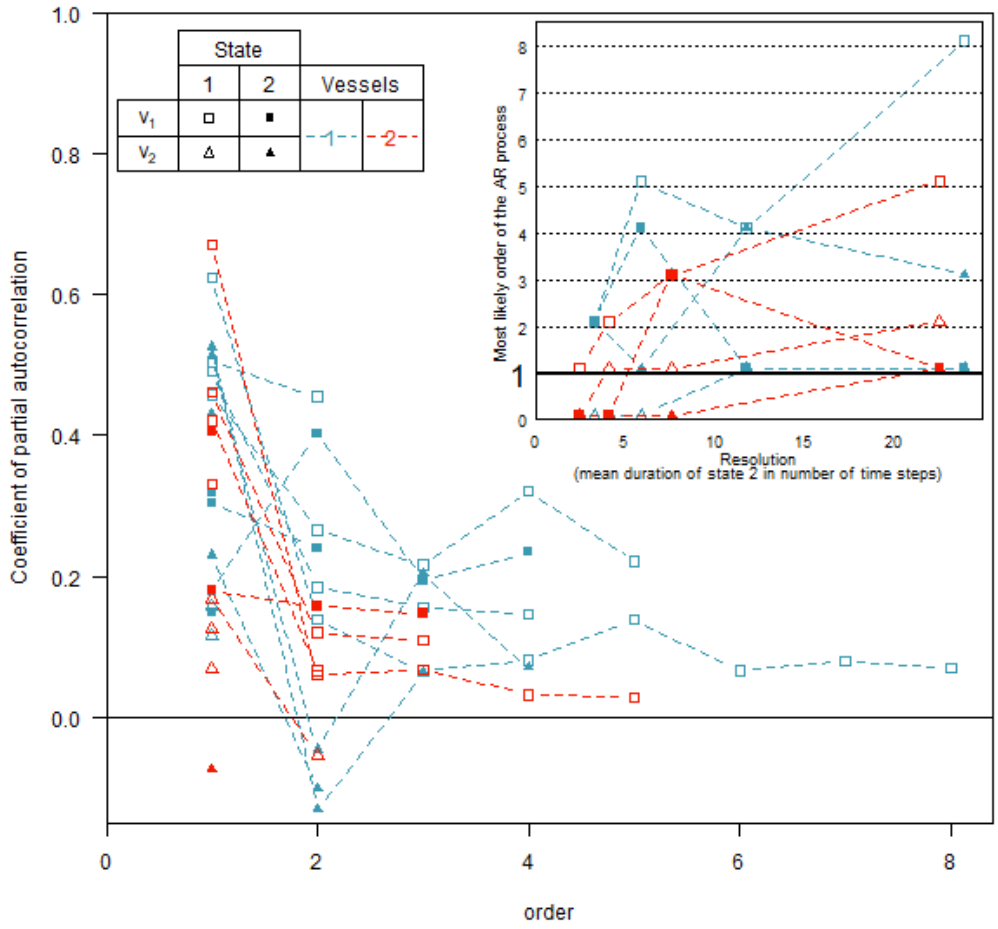

Figure 14. Partial auto-correlations of state-dependent variables. For each variable, vessel, resolution, and state, the partial auto-correlation coefficients are represented up to the point when they are no longer statistically different from 0. The inner panel summarizes the order obtained in each case with a reference being order 1.

Discussion

Scope of the benchmarking

The current study's benchmark includes four model types (some with nesting properties) and two sets of contrasted settings. We chose to focus on these restricted configurations rather than following an experimental design crossing parameter values over a wider range. such a comprehensive experimental design would have required a significant amount of computing time. Instead, we defined the parameters of the four types of models using two sets of real fishing vessel trajectories that provided ground truth on the sequence of states. While this likely limits the overall generalizability of the results, it allows for useful recommendations for various real-world situations involving other natural foragers than fishers. According to Nettle et al., 2013, the growing interest in human behavioral ecology presents a significant opportunity to bridge the gap between natural and social sciences. Regarding fisheries science, Bertrand et al., 2007 show that fishermen follow similar patterns as natural top predators. In this context, specifying the models using annotated vessel trajectories places the study in a broader, more relevant ecological context. Without being too general, the scope of this work is quite broad. The model parameters are sufficiently contrasted to distinguish first-order

auto-regressive processes from their AR0 counterparts and HSMMs from their HMM counterparts. So, the model parameters cause a significant contrast between the four model types.

The four model structures investigated in the experiments are reasonably simple (two hidden states with two independent observed variables). Most models in movement ecology (including fisheries) adopt simple modeling frameworks (e.g. Bez et al., 2011; Langrock et al., 2012; Vermard et al., 2010) that are still recommended in more recent studies (Auger-Méthé, Newman, et al., 2021; McClintock, 2021). The number of states and the identification of a specific state among them are sometimes the primary goals of using state space models, particularly in behavioral ecology. The goal of this analysis is not to infer the possibility of correctly estimating the *number* of states but rather to quantify the state decoding accuracy of hidden Markov models under controlled conditions. If you are interested in this issue, we recommend reading the pragmatical approach proposed by Pohle et al., 2017. Others sources of possible biases coming from temporal misalignment between model and observations are also not considered (Glennie et al., 2023). Although not shown, two intriguing side outputs are worth noting. First, the Estimation-Maximization (EM) algorithm produced reliable estimates of maximum likelihood, i.e., not impacted by changes in initialization values. Second, the most probable state sequences (Viterbi) were equal to marginal estimates, implying that the Viterbi algorithm is unnecessary in these cases.

No best model in model skill assessment conditions

The best model performances are obtained from model skill assessments, which occur when the simulation and estimation models are identical. This is consistent with the fact that, given a fixed number of observations, one cannot expect to achieve higher accuracy when the data and model are no longer consistent. Given favorable inference situations (i.e., 250 state alternations), all four model types, including the most complex (e.g., HSMM-AR1), achieve similar state decoding accuracies and very good parameter estimations. This is not an exceptional situation; one of our two case studies shows the same amount of state alternation. This means that there is no optimal model type under MSA conditions. Model parameters and settings (including resolution) control performance rather than the model structure itself. The following examples illustrate this statement. In both settings, the variances of the sojourn times are greater in HMM formulation than in HSMM formulation (Figure 7).

The only distinction between the two is that, for Settings 2, the variance of the sojourn time is an order of magnitude greater in State 2 than in State 1, while the proportion of time spent in State 2 is quite small (π_2=0.15, π_1=0.85). This could have been detrimental to the fittings with Settings 2. However, we observe the opposite. The difference in performances is driven by the two metrics (d_V and d_{AR}) that are clearly in favor of Settings 2 (Figure 5) and that explain strongly the shapes and levels of the different accuracies (Figure 9). Regardless of model type, the characteristics of the observation layer (more favorable for Settings 2 than for Settings 1) explain the performance better than those of the hidden layer. Under MSA conditions, state-space models consistently outperform mixed models (Figure 9). This finding suggests that the structures introduced by temporal state processes facilitate state estimation, at least as long as the estimation model agrees with the simulation model. Consider applying this result to real-world situations: in that case, the ability to select the true temporal process will have a greater impact on state inference quality than the change state assumption.

Auto-correlation deteriorates the state decoding accuracy

The numerical experiments revealed that the deterioration in accuracy is primarily due to auto-correlation in the observation layers rather than the hidden layer's Markovian structure. Although d_T (hidden layer) and d_{AR} (observation layer) are co-linear in the settings with potential confounding effects, the current work demonstrates that the modeling choices related to the observation layer are of primary importance for state

decoding performances. This finding is supported by both simulation-estimation experiments and real-data estimation.

Simulation-estimation experiments

The HSMM$_{s\mathcal{NB}}$ - AR1 model structure is the most robust to discrepancies between model hypotheses and data characteristics when retrieving hidden states, according to model robustness assessment (MRA). While the accuracy in state estimation deteriorates in 90% of cases when using a model structure different from the one used to simulate data, the relative loss is small (0.92% on average). Imposing no auto-correlation in the estimating model while allowing some for state-dependent variables has a negative impact on state decoding performance. The loss of accuracy is well explained by d_{AR}, which is the distance between AR1 and AR0 formulations. This is not true for the hidden layer, where imposing an HMM structure while the state process is semi-Markov has no negative impact on the state decoding performance. This finding has direct practical implications. It demonstrates that it is preferable to concentrate on the hypothesis choices of the observation layer rather than the hidden layer. In some ways, this is reassuring, as only the observation layer contains empirical data. Even though, in theory, the time dependence of the observed processes or the distinction of the PDFs should be evaluated conditionally on the hidden states (which is impossible in practice because the states are, by definition, hidden), we recommend conducting exploratory analyses of the observed process data to define auto-correlation hypotheses.

Real cases

The analysis of model performance on real cases is consistent with the simulation results, indicating that discrepancies between the properties of the observations and the model's hidden layer assumptions do not affect performance. Both data sets demonstrate a visual deviation from the various model assumptions. However, even though the differences in residence time for Vessel 2 are greater, the model's performance improves. The most robust model (HSMM$_{s\mathcal{NB}}$ - AR1) results in lower state decoding performance for Vessel 1 and higher performance for Vessel 2. Interestingly, Vessel 2's performances do not deteriorate strongly with the resolution, whereas Vessel 1's do (Figure 9). The two primary differences in favor of Vessel 2 concern the speed variables:

- The discriminating power between the speed distributions measured through the integrated metric d_V is low, respectively large, for Vessel 1 and Vessel 2 (Figure 13). This contributes to the low accuracies obtained for Vessel 1.
- In most cases, the partial autocorrelation of observed speed is greater than one (the model's assumed autocorrelation), particularly for Vessel 1. Not only are the auto-correlations of order 1 larger, but the most likely order of the auto-regressive process modeling the speed variable is greater than 1.

Beyond these considerations about the value of coefficients of auto-correlation and the orders of the auto-regressive processes, the observations may not be consistent with an auto-regressive process at all. In particular, auto-regressive processes assume that the relationship between successive observations is linear, which may not be the case. As an example, we provide the cross-plots between successive speed components (Additional Figure 15), which violate linear relationships over time. To form the correct hypotheses about the observation process, start by ensuring that the data are consistent with an auto-regressive process. The violation of this assumption corresponds to a structural error process that is rarely evaluated. Before discussing the order of the AR, consider the more fundamental choice of using an AR formulation for the observation layer (Gloaguen et al., 2015; Joo et al., 2013; Morales et al., 2004). Second, within the context of auto-regressive processes, the data may not correspond to a zero or first AR order (AR0 or AR1). This second type of violation is caused by mis-specification and can be more easily controlled by evaluating the model's goodness of fit.

Increasing the resolution: a tricky solution?

As expected, model characteristics are linked to data resolution. This affects not only the averages of the sojourn times, which are directly impacted because they are expressed in a series of time steps but also their distributions. This also affects the variability of the speeds (both persistent and rotational), their time dependencies, and the overlap of speed PDFs between states. The multidimensional nature of data resolution precludes statistical investigation of its effect on estimation performance, all other things being equal. Several authors (e.g., McClintock, King, et al., 2012; Vermard et al., 2010) have pointed out that time steps must be chosen to match the scale at which behavioral decisions are made using expert knowledge. Katara and Silva Katara and Silva, 2017 examined the mismatch between the temporal resolution of position records and the timescale of fishing activity. The authors investigated the effect of time resolution (from 10 minutes to 2 hours) on bias and errors in identifying fishing operations for vessels (Portuguese purse seiners) with characteristics similar to those of vessel 2. The results (7% of missed fishing sets with 2h time interval compared to 10 minutes) align with the estimated increase in accuracy with time scale resolution for Vessel 2.

Similar results were obtained for other vessels, emphasizing the importance of high-resolution time data Edelhoff et al., 2016; Mendo et al., 2019. Interestingly, when using data that were consistent with a given model's assumptions, such as in simulation-estimation experiments, increasing the resolution of simulated data resulted in higher accuracy. This was less obvious for real data, where the results differed between the two case studies and were not linear with data resolution. There is a contradiction between the need for high-frequency data to track short-term states or to get as close as possible to the tracked individual's actual behavior and the ability to model this behavior using a random decision. If we suppose that memory and/or inertia play a role in behavioral changes at high frequencies, in this case, it is clear that a state is dependent not only on previous states but also on multiple previous ones. As a result, a compromise must be made between the pursuit of better statistical conditions and a realism threshold to achieve an optimal level of data resolution.

In conclusion, in silico, when data characteristics are controlled by and consistent with model characteristics, simulation-estimation experiments show that a change in Markov property for the estimation model has no negative impact on state decoding performance unless one forces the state-dependent variable to be AR0 when they are indeed AR1. This aligns with the fact that HSMM$_{s\mathcal{NB}}$ and AR1 formulations are generalizations of the usual HMM and AR0 formulations. This study also shows that model performance is more sensitive to the auto-regressive characteristics of the state-dependent variables than to the Markov properties of the hidden states. This is somewhat satisfactory because, in practice, one may have access to some aspects of the observation layer but not to the hidden, unobserved layer. This allows for empirical controls on what appear to be key model assumptions in terms of state decoding performance. However, in real cases, the robustness of HSMM$_{s\mathcal{NB}}$ - AR1 appears low, with instances where its state decoding performances are worse than simple filtering. The confusion between the distributions of state-dependent variables in different states, whether observable (usually not possible) or more realistically speculated, is the most important factor influencing state decoding performance in real-world scenarios. Blind applications of state-space models in the absence of prior knowledge about the processes lead to radical errors and poor state-decoding performance. Given that HMMs are a popular choice in ecology due to their ease of use, we believe there is a need for serious consideration of the consequences of the numerous potential deviations between data and model assumptions.

Appendices

Supplementary figure

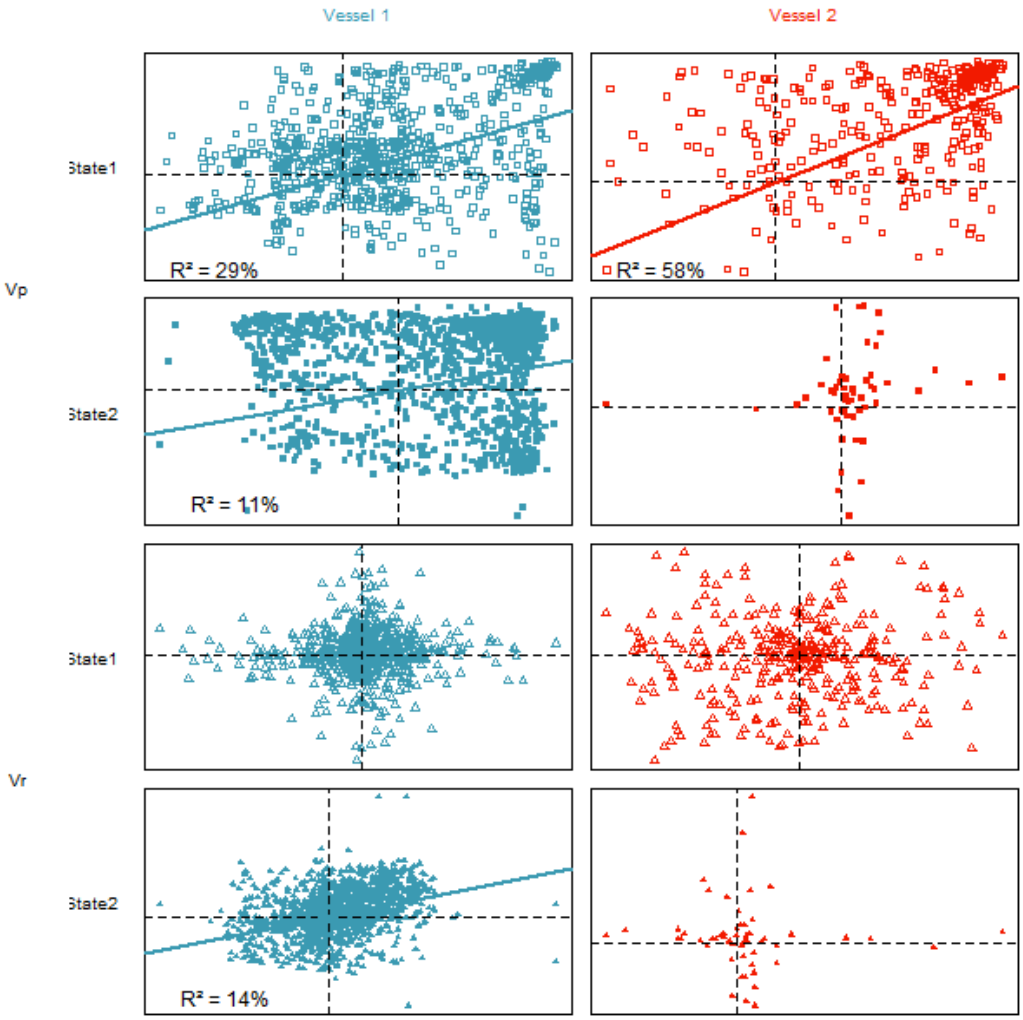

Figure 15. SUPPLEMENTARY. Delayed correlation plots for Vessel 1 (left, blue panels) and Vessel 2 (right, red panels). Persistent speed is in squares, and rotation speed is in triangles. State 1 is in empty symbols, and State 2 is in plain. Linear regressions for which p-value < 5% are represented.

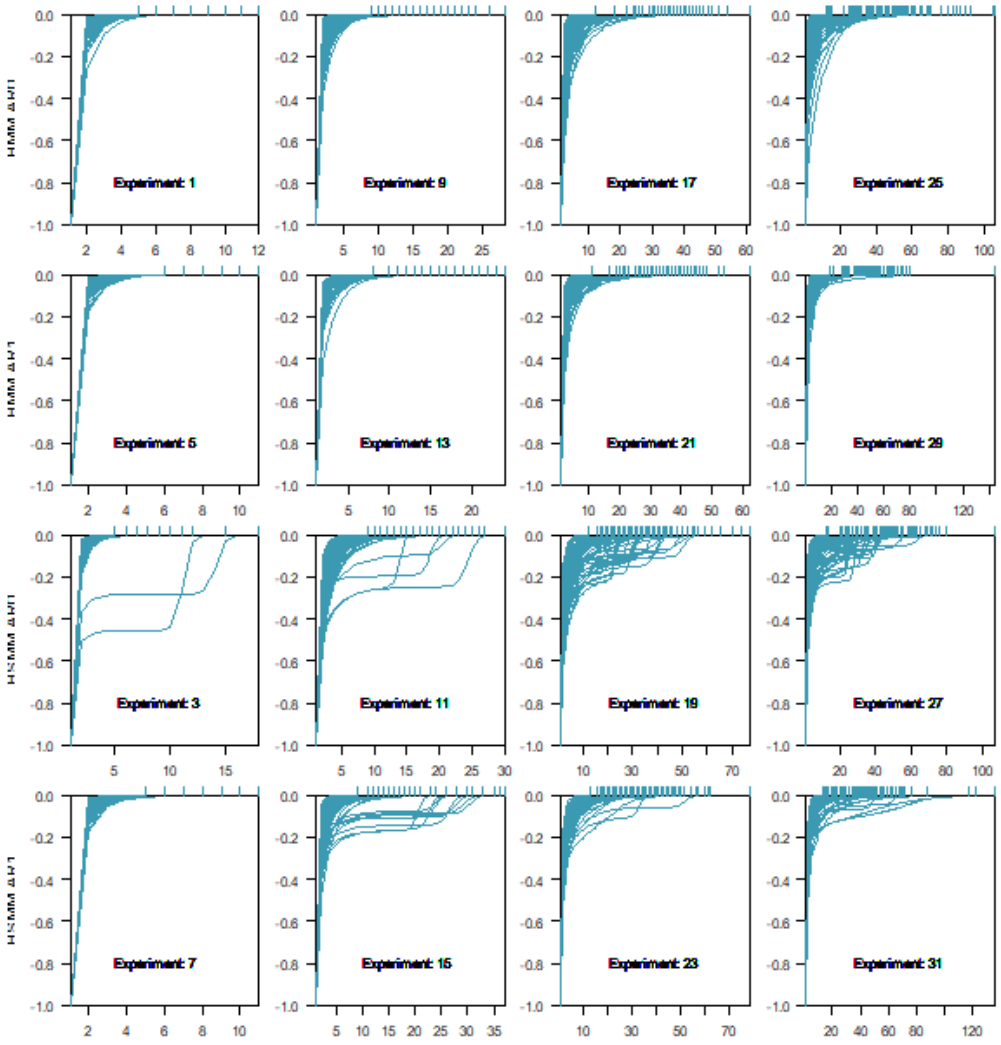

Figure 16. SUPPLEMENTARY. Settings 1. Evolution of the log-likelihood of each EM under MSA conditions (estimation model = simulation model. Rows correspond to the different model types. Columns correspond to the different resolutions. On each panel, the evolution of the log-likelihood during each of the 100 estimations is represented. The log-likelihood is normalized between the value for the initialization and the value obtained after convergence. The x-axis represents the number of iterations (different for each experiment). Tick marks represent the number of iterations reached at each estimation. When the EM reaches the maximum possible number of iterations, the output is represented by a black curve (which never happens as the numerical settings have been chosen so that this does not happen).

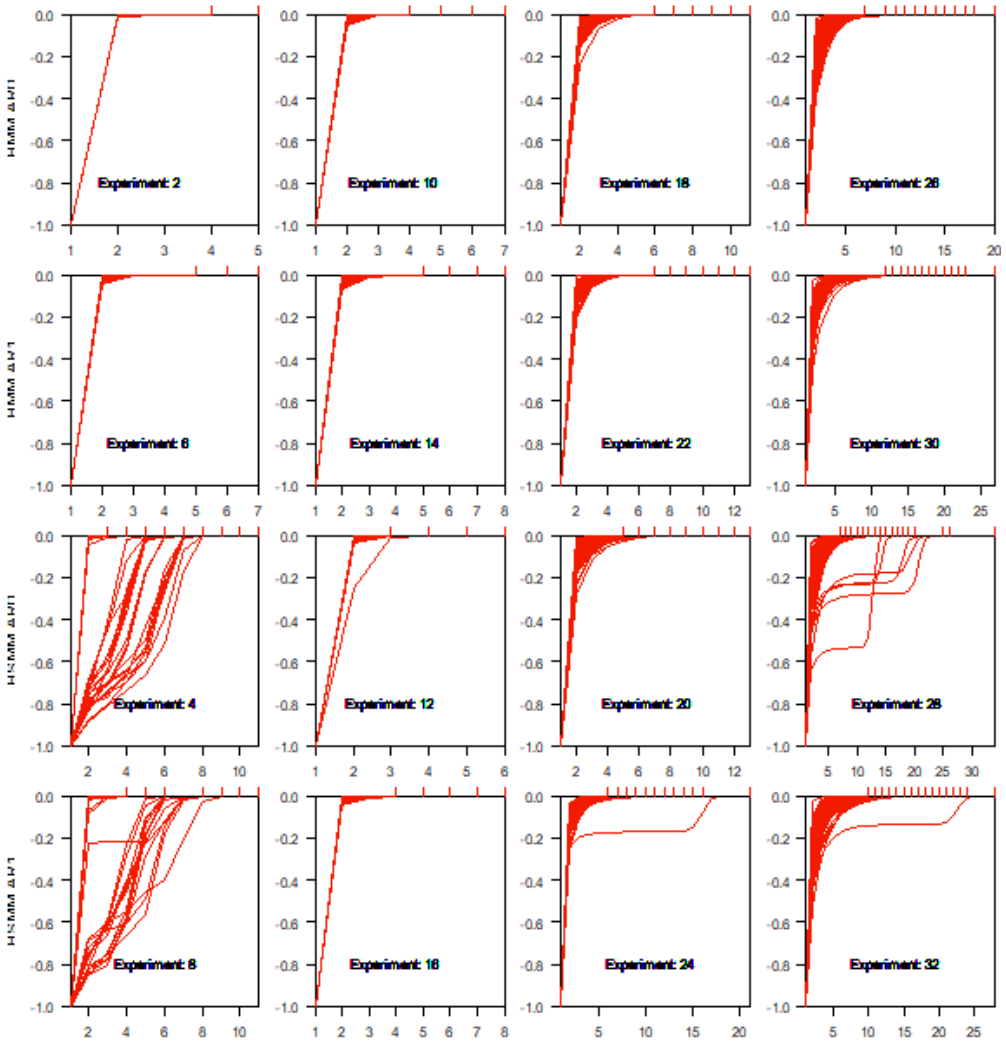

Figure 17. SUPPLEMENTARY. Settings 2. Same caption as before.

Additional information: On two occasions, the estimation of the shift for the shifted negative binomial PMFs causes a decreasing evolution of the likelihood. In these cases, we changed the seed to recover the normal behavior of the convergence. These cases correspond to experiment n^o 4 - simulation n^o72, and to experiment n^o 12 - simulation n^o73. See the code to get details (https://github.com/nicobez/H-S-MM_MSA).

Acknowledgements

The authors acknowledge the Pôle de Calcul et de Données Marines (PCDM) for providing DATARMOR storage, data access, computational resources, visualization, web services, consultation, and support services. URL: http://www.ifremer.fr/pcdm.

Several colleagues have contributed to discussions and meetings without being formally co-authors. Their inputs are worth to be acknowledged: Pascal MONESTIEZ, David NERINI, Hélène De PONTUAL, Youen VERMARD.

Fundings

The participation in the various workshops within which this work was discussed was financed by the INRAE network Path Tools and analysIS - PathTIS.

Conflict of interest disclosure

The authors declare that they comply with the PCI rule of having no financial conflicts of interest in relation to the content of the article.

Data, script, code, and supplementary information availability

Model simulations and states' estimations are done within R version 4.3.1 "Beagle Scouts" (R Core Team, 2022) using the package mhsmm (O'Connell and Højsgaard, 2011), which allows performing EM inferences with several realizations of the same underlying model. Modifications of the mhsmm package were, however, required. We extended the package:

- to the case of AR1 for the state-dependent variables that were not in the package,
- to the case of shifted negative binomial PMF for the sojourn time that was not fully implemented,
- to force independence between the state-dependent variables during the inference process (by default, a correlation is re-estimated during the EM procedure).

When considering 250 alternations of State 1 and 2, the 32 MSA experiments corresponding to the 2 settings $\times$ 4 degradations $\times$ 4 model types are fully reproducible from https://doi.org/10.5281/zenodo.10678877. Tolerances for stopping the iterative search of the maximum likelihood in the EM algorithm are adapted to avoid very long searches associated with teeny marginal improvements of the likelihood (the iterative procedure stops when the gain in log-likelihood is smaller than the tolerance). The tolerance is reduced to 10^{-04} for HMM, and to 10^{-03} for HSMM when the default values were 10^{-08} for HMM, and 10^{-04} for HSMM. Meanwhile, the maximum number of iterations is also adapted to ensure that the convergence is reached, at least with regard to the chosen tolerance. The maximum number of iterations is reduced to 200 for HMM (instead of 1000 by default) but increased to 150 for HSMM (instead of 100 by default). With these numerical settings, convergence is systematically reached before the maximum possible number of iterations (Figures 16 and 17).

With these numerical settings, the full MSA experiments last 10 hours and 20 minutes on an i7 PC which probably ruins their fluid and rapid full reproduction. This certainly does for the MRA procedures that are

3 times longer. Thus, the MRA simulation-estimation experiments have been performed remotely using the computational resources of a super-computer (DATARMOR, http://www.ifremer.fr/pcdm). All the simulation-estimation outputs have been stored and the paper is reproducible based on this stored and open-access information. The code to re-run all the MRA simulations is, however, available on demand. After the upload of all the outputs of the simulation-estimation experiments, all the analyses and figures are reproducible from https://doi.org/10.5281/zenodo.10679448. For anonymity concerns, the coordinates of the trajectories are centered and standardized.

References

Allen J and P Somerfield (2009). A multivariate approach to model skill assessment. *Journal of Marine Systems* 76, 83–94. ISSN: 0924-7963. https://doi.org/10.1016/j.jmarsys.2008.05.009.

Auger-Méthé M, CCS Clair, MA Lewis, and AE Derocher (Aug. 2011). Sampling rate and misidentification of Lévy and non-Lévy movement paths: comment. en. *Ecology* 92, 1699–1701. ISSN: 0012-9658. https://doi.org/10.1890/10-1704.1.

Auger-Méthé M, K Newman, D Cole, F Empacher, R Gryba, AA King, V Leos-Barajas, J Mills Flemming, A Nielsen, G Petris, and L Thomas (2021). A guide to state–space modeling of ecological time series. en. *Ecological Monographs* 91, e01470. ISSN: 1557-7015. https://doi.org/10.1002/ecm.1470.

Avgar T, R Deardon, and JM Fryxell (Feb. 2013). An empirically parameterized individual based model of animal movement, perception, and memory. *Ecological Modelling* 251, 158–172. ISSN: 03043800. https://doi.org/10.1016/j.ecolmodel.2012.12.002.

Bertrand S, A Bertrand, R Guevara-Carrasco, and F Gerlotto (2007). Scale-Invariant Movements of Fishermen: The Same Foraging Strategy as Natural Predators. en. *Ecological Applications* 17, 331–337. ISSN: 1939-5582. https://doi.org/10.1890/06-0303.

Bez N, E Walker, D Gaertner, J Rivoirard, and P Gaspar (Nov. 2011). Fishing activity of tuna purse seiners estimated from vessel monitoring system (VMS) data. en. *Canadian Journal of Fisheries and Aquatic Sciences* 68. Ed. by Walters C, 1998–2010. ISSN: 0706-652X, 1205-7533. https://doi.org/10.1139/f2011-114.

Dempster AP, NM Laird, and DB Rubin (Sept. 1977). Maximum Likelihood from Incomplete Data Via the *EM* Algorithm. en. *Journal of the Royal Statistical Society: Series B (Methodological)* 39, 1–22. ISSN: 00359246. https://doi.org/10.1111/j.2517-6161.1977.tb01600.x.

Edelhoff H, J Signer, and N Balkenhol (Dec. 2016). Path segmentation for beginners: an overview of current methods for detecting changes in animal movement patterns. en. *Movement Ecology* 4, 21. ISSN: 2051-3933. https://doi.org/10.1186/s40462-016-0086-5.

Gerritsen H and C Lordan (Jan. 2011). Integrating vessel monitoring systems (VMS) data with daily catch data from logbooks to explore the spatial distribution of catch and effort at high resolution. en. *ICES Journal of Marine Science* 68, 245–252. ISSN: 1095-9289, 1054-3139. https://doi.org/10.1093/icesjms/fsq137.

Glennie R, T Adam, V Leos-Barajas, T Michelot, T Photopoulou, and BT McClintock (2023). Hidden Markov models: Pitfalls and opportunities in ecology. en. *Methods in Ecology and Evolution* 14, 43–56. ISSN: 2041-210X. https://doi.org/10.1111/2041-210X.13801.

Gloaguen P, S Mahevas, E Rivot, M Woillez, J Guitton, Y Vermard, and MP Etienne (Feb. 2015). An autoregressive model to describe fishing vessel movement and activity. *Environmetrics* 26. WOS:000347896200003, 17–28. ISSN: 1180-4009. https://doi.org/10.1002/env.2319.

Guédon Y (2003). Estimating Hidden Semi-Markov Chains from Discrete Sequences. en. *Journal of Computational and Graphical Statistics* 12, 604–639.

Hellinger E (July 1909). Neue Begründung der Theorie quadratischer Formen von unendlichvielen Veränderlichen. en. *Journal für die reine und angewandte Mathematik* 1909, 210–271. ISSN: 1435-5345, 0075-4102. https://doi.org/10.1515/crll.1909.136.210.

Johnson MJ and AS Willsky (2013). Bayesian Nonparametric Hidden Semi-Markov Models. *Journal of Machine Learning Research* 14, 673–701.

Jonsen I, M Basson, S Bestley, M Bravington, T Patterson, M Pedersen, R Thomson, U Thygesen, and S Wotherspoon (Apr. 2013). State-space models for bio-loggers: A methodological road map. en. *Deep Sea Research Part II: Topical Studies in Oceanography* 88-89, 34–46. ISSN: 09670645. https://doi.org/10.1016/j.dsr2.2012.07.008.

Joo R, S Bertrand, J Tam, and R Fablet (2013). Hidden markov models: the best models for forager movements? *PloS one* 8, e71246.

Katara I and A Silva (Apr. 2017). Mismatch between VMS data temporal resolution and fishing activity time scales. en. *Fisheries Research* 188, 1–5. ISSN: 01657836. https://doi.org/10.1016/j.fishres.2016.11.023.

Langrock R, R King, J Matthiopoulos, L Thomas, D Fortin, and JM Morales (2012). Flexible and practical modeling of animal telemetry data: hidden Markov models and extensions. *Ecology* 93, 2336–2342. ISSN: 0012-9658. https://doi.org/10.1890/11-2241.1.

Lantuéjoul C (1991). Ergodicity and integral range. *Journal of Microscopy* 161, 387–403.

Lynch DR, DJJ McGillicuddy, and FE Werner (Feb. 2009). Skill assessment for coupled biological/physical models of marine systems. en. *Journal of Marine Systems* 76, 1–3. ISSN: 0924-7963. https://doi.org/10.1016/j.jmarsys.2008.05.002.

McClintock BT (2021). Worth the effort? A practical examination of random effects in hidden Markov models for animal telemetry data. en. *Methods in Ecology and Evolution* 12, 1475–1497. ISSN: 2041-210X. https://doi.org/10.1111/2041-210X.13619.

McClintock BT, R King, L Thomas, J Matthiopoulos, BJ McConnell, and JM Morales (2012). A general discrete-time modeling framework for animal movement using multistate random walks. *Ecological Monographs* 82, 335–349.

McClintock BT, R Langrock, O Gimenez, E Cam, DL Borchers, R Glennie, and TA Patterson (Dec. 2020). Uncovering ecological state dynamics with hidden Markov models. en. *Ecology Letters* 23. Ed. by Coulson T, 1878–1903. ISSN: 1461-023X, 1461-0248. https://doi.org/10.1111/ele.13610.

Mendo T, S Smout, T Russo, L D'Andrea, and M James (Dec. 2019). Effect of temporal and spatial resolution on identification of fishing activities in small-scale fisheries using pots and traps. en. *ICES Journal of Marine Science* 76. Ed. by Maravelias C, 1601–1609. ISSN: 1054-3139, 1095-9289. https://doi.org/10.1093/icesjms/fsz073.

Morales JM, DT Haydon, J Frair, KE Holsinger, and JM Fryxell (Sept. 2004). Extracting more out of relocation data: building movement models as mixtures of random walks. en. *Ecology* 85, 2436–2445. ISSN: 0012-9658. https://doi.org/10.1890/03-0269.

Nettle D, MA Gibson, DW Lawson, and R Sear (2013). Human behavioral ecology: current research and future prospects. en. *Behavioral Ecology* 24, 1031–1040. ISSN: 1465-7279, 1045-2249. https://doi.org/10.1093/beheco/ars222.

O'Connell J and S Højsgaard (2011). Hidden Semi Markov Models for Multiple Observation Sequences: The **mhsmm** Package for *R*. en. *Journal of Statistical Software* 39. ISSN: 1548-7660. https://doi.org/10.18637/jss.v039.i04.

Owen-Smith N, JM Fryxell, and EH Merrill (2010). Foraging theory upscaled: the behavioural ecology of herbivore movement. *Philosophical Transactions of the Royal Society B: Biological Sciences* 365, 2267–2278. https://doi.org/10.1098/rstb.2010.0095.

Pohle J, R Langrock, FM van Beest, and NM Schmidt (Sept. 2017). Selecting the Number of States in Hidden Markov Models: Pragmatic Solutions Illustrated Using Animal Movement. en. *Journal of Agricultural, Biological and Environmental Statistics* 22, 270–293. ISSN: 1085-7117, 1537-2693. https://doi.org/10.1007/s13253-017-0283-8.

Poritz A (1982). Linear predictive hidden Markov models and the speech signal. *Proc.ICASSP '82. pp.1291-1294, Paris, France, May 1982*.

R Core Team (2022). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. Vienna, Austria.

Rabiner L (Feb. 1989). A tutorial on hidden Markov models and selected applications in speech recognition. *Proceedings of the IEEE* 77, 257–286. ISSN: 00189219. https://doi.org/10.1109/5.18626.

Schwager M, DM Anderson, Z Butler, and D Rus (Mar. 2007). Robust classification of animal tracking data. en. *Computers and Electronics in Agriculture* 56, 46–59. ISSN: 0168-1699. https://doi.org/10.1016/j.compag.2007.01.002.

Sonia Malefaki ST and N Limnios (2010). An EM and a Stochastic Version of the EM Algorithm for Nonparametric Hidden Semi-Markov Models. *Communications in Statistics - Simulation and Computation* 39, 240–261. https://doi.org/10.1080/03610910903411185.

Sur M, T Suffredini, SM Wessells, PH Bloom, M Lanzone, S Blackshire, S Sridhar, and T Katzner (2017). Improved supervised classification of accelerometry data to distinguish behaviors of soaring birds. en. *PLOS ONE* 12, e0174785. https://doi.org/10.1371/journal.pone.0174785.

Thums M, CJ Bradshaw, and MA Hindell (Dec. 2008). A validated approach for supervised dive classification in diving vertebrates. en. *Journal of Experimental Marine Biology and Ecology* 363, 75–83. ISSN: 0022-0981. https://doi.org/10.1016/j.jembe.2008.06.024.

Vermard Y, E Rivot, S Mahévas, P Marchal, and D Gascuel (2010). Identifying fishing trip behaviour and estimating fishing effort from VMS data using Bayesian Hidden Markov Models. *Ecological Modelling* 221, 1757–1769. ISSN: 0304-3800.

Yaglom A (1987). *Correlation Theory of Stationary and Related Random Functions*. en.

Zucchini W and IL MacDonald (2009). *Hidden Markov models for time series: an introduction using R*. Monographs on statistics and applied probability 110. OCLC: ocn144565537. Boca Raton: CRC Press. ISBN: 9781584885733.

Table 1. Definition of the notations

	Notation	Meaning
	t	Discrete time step
Hidden layer	s_t , S_t	State at time step t (lower case: observed; upper case:random variable)
	N_S	Number of possible states
	t_s , T_s	Sojourn time in state s (lower case: observed; upper case:random variable)
	$p_{s,s}$	For HMM, probability to stay in state s when moving from t to $t+1$ so that $T_s \sim geom(1-p_{s,s})$
	N_{pS}	Number of parameters of the PMFs of the sojourn time
	$shift_s$	Shift for the PMF of T_s for HSMM
	(n_s, p_s)	Parameters of the negative binomial of the PMF of T_s for HSMM so that $T_s \sim shift_s + \mathcal{NB}(n_s, p_s)$
	N_H	Number of parameters for the hidden layer
	d_{T_s}	Hellinger distance between the geometric and the shifted negative binomial PMFs of T_s
	d_T	Mean Hellinger distance between the geometric and the shifted negative binomial PMFs over all possible states
	π_s	Probability to be in state s
Observation layer	N_V	Number of state-dependent variables
	v_i , V_i	State-dependent variable number i (lower case: observed; upper case:random variable)
	(V_p, V_r)	Persistent and rotation speeds (V_1, V_2)
	$V_{i,t}$	Temporal process for state-dependent variables
	$\mathcal{N}(\mu_{i,s}, \sigma_{i,s}) = \mathcal{N}_{i,s}$	Uni-variate Gaussian PDFs of the state-dependent variable i for state s
	$\mu_{i,s}$	Expected value of $V_{i,s}$
	$\sigma_{i,s}$	Standard deviation of $V_{i,s}$
	N_{p0}	Number of parameters of the PDFs of the state-dependent variables
	N_O	Number of parameters for the observation layer
	d_V	Discrimination between PDFs
	$\tau_{i,s}$	Coefficient of correlation for auto-regressive processes of order 1 for state-dependent variable i in state s
	$\eta_{i,s}$	$\mu = \frac{eta}{1-tau}$
	A	Integral range of AR1 process
	$d_{AR_{i,s}}$	Difference between AR0 and AR1 formulations for state-dependent variable i and state s
	d_{AR_s}	Difference between AR0 and AR1 formulations for state-dependent variables in state s
	d_{AR}	Difference between AR0 and AR1 formulations

Table 2. Statistics for Vessel 1. For the sojourn time in the various states, the parameters of the PMFs in the model settings (e.g. $p_{s,s}, n_s, p_s$) are deduced from them according to the equations provided in the method section. The statistics for the sojourn time are provided in the number of time steps. For the speed variables, they are expressed in knots.

	Resolution 15 min (N=9506)		Resolution 30 min (N=4748)		Resolution 60 min (N=2370)		Resolution 120 min (N=1171)	
	T							
	T_1	T_2	T_1	T_2	T_1	T_2	T_1	T_2
$E[T_s]$	11	24	5.7	12	2.9	6.1	1.6	3.5
$\sigma[T_s]$	7.4	11	3.9	5.7	2	3	1.1	2.1
$shift_s$	2	1	1	1	1	1	1	1
	$V_1 = V_p$							
	$V_{1,1}$	$V_{1,2}$	$V_{1,1}$	$V_{1,2}$	$V_{1,1}$	$V_{1,2}$	$V_{1,1}$	$V_{1,2}$
$\mu_{1,s}$	4.5	3	3.8	2.4	2.7	1.4	1.8	0.3
$\sigma_{1,s}$	4.6	1.6	4.9	2.0	5.1	2.4	4.7	2.2
$\tau_{1,s}$	0.60	0.34	0.46	0.15	0.43	0.17	0.41	0.30
	$V_2 = V_r$							
	$V_{2,1}$	$V_{2,2}$	$V_{2,1}$	$V_{2,2}$	$V_{2,1}$	$V_{2,2}$	$V_{2,1}$	$V_{2,2}$
$\mu_{2,s}$	0.07	0.13	0.11	0.24	0.08	0.22	0.12	0.009
$\sigma_{2,s}$	1.7	1.4	2	1.5	2.2	1.4	2.3	1.3
$\tau_{2,s}$	0.15	0.52	0.11	0.50	0.07	0.43	-0.008	0.27

Table 3. Statistics for Vessel 2. For the sojourn time in the various states, the parameters of the PMFs in the model settings (e.g. $p_{s,s}, n_s, p_s$) are deduced from them according to the equations provided in the method section. The statistics for the sojourn time are provided in the number of time steps. For the speed variables, they are expressed in knots.

	Resolution 5 min (N=6181)		Resolution 15 min (N=2011)		Resolution 30 min (N=968)		Resolution 60 min (N=446)	
	T							
	T_1	T_2	T_1	T_2	T_1	T_2	T_1	T_2
$E[T_s]$	104	22	41	7.4	20	4.4	11	2.3
$\sigma[T_s]$	112	4.4	45	1.9	22	1.8	12	1.2
$shift_s$	4	15	1	4	1	1	1	1
	$V_1 = V_p$							
	$V_{1,1}$	$V_{1,2}$	$V_{1,1}$	$V_{1,2}$	$V_{1,1}$	$V_{1,2}$	$V_{1,1}$	$V_{1,2}$
$\mu_{1,s}$	9.3	0.46	8.3	0.43	7.3	0.41	6.5	0.32
$\sigma_{1,s}$	4.7	1.2	5.2	0.9	5.5	1.3	5.5	1.6
$\tau_{1,s}$	0.70	0.56	0.49	0.27	0.44	0.12	0.35	0.20
	$V_2 = V_r$							
	$V_{2,1}$	$V_{2,2}$	$V_{2,1}$	$V_{2,2}$	$V_{2,1}$	$V_{2,2}$	$V_{2,1}$	$V_{2,2}$
$\mu_{2,s}$	0.13	-0.08	-0.03	0.11	-0.24	0.20	-0.31	0.20
$\sigma_{2,s}$	2.9	0.80	3.8	0.80	4.3	1.3	4.6	1.6
$\tau_{2,s}$	0.17	-0.10	0.12	-0.04	0.07	-0.09	0.04	-0.07