Answer to reviewers

Bez et al.

November 7, 2024

The authors want to underline the very positive kindness in reviewers comments, requests with personal suggestions and their detailed analysis of the manuscript. Reading their feed backs and their comments has been a pleasure and a relevant material to improve the manuscript. Would all the reviews be done with the same good-will and rigour.
We addressed all but some comments (with justifications in orange for the few points we did not addressed). All the points raised by the reviewers get answers and all the modifications in the MS are tractable (except the table of notation that has been strongly modified). By taking these remarks into account, we truly believe that the MS was improved. This phase of the review process has been pleasant and stimulating.

Reviewer 1

Review

The authors propose a study of the impact of model mis-specification for models on the family of HMM and HSMM. Mis-specification can be at the level of the hidden chain (Markov versus semi-Markov) or at the level of the observed chain (AR0 versus AR1). The study is made in the context of data from fishery vessel movements. The impact of model mis-specification is assessed on the restored hidden chain (decoding task), which I find very relevant since in many applications we are more interested by decoding quality rather than by precise parameters estimation. The main conclusion of the study is that choosing the wrong AR model at the observed sequence level has more impact that choosing the wrong model on the hidden chain. This work addresses a very interesting topic for statisticians and ecological modelers. As underlined by the two reviewers it is very clearing presented. However, they have made comments that require answers from the authors and clarification in the manuscript. I also have some remarks and questions listed below.

P.1.1 *One main conclusion is that 'imposing a Markov structure while the state process is semi-Markov does not impair the state decoding performance'. Actually, this result is obtained for a particular semi-Markov model, with Negative Binomial sojourn time distribution. I understand the reasons for restricting the analysis to this distribution but then, the conclusion cannot be so general. Maybe with another sojourn time distribution, the impact on decoding would be more important.*

We agree. The statement is now more detailed. See Line 15.

P.1.2 *Figure 1. There are some approximations in the legend text. The legend starts with 'Directed acyclic graph for HMM and HSMM, . . . '. I agree with the term in the case of HMM. But not in the HSMM case. In a DAG representation of a Bayesian network, the nodes of the graph are the model variables and in the classical HSMM representation variables are jump time, sojourn duration or date of jump, and observations indexed on calendar time. Since the value of each sojourn time is not known in advance it is not possible to draw the arrows from hidden states towards observations. Still about the legend, the arrows in a DAG can enable to recover conditional independencies. But an absence of arrow does not indicate that the two variables are conditionally independent (it would depend on which variable is the conditionning). This can be seen in the case of a V-structure.*

These remarks have been considered. We agree with the recommender that the nodes are the variables and that, for HSMM, the jump time should appear in the graph. However, the state itself

is also a variable. While in the present 2-states HSMM, the state is quasi-deterministic, i.e. leaving state 1 we have no other choice than jumping into state 2, and vice-versa, the state should appear in the figure. This also means that the values taken by the state variable should not appear on the figure but only the variable.
We tried to improve the figure to account for all this :

- we modified the legend text accordingly. For simplicity, we no longer speak about conditional independence
- we remove the black and white indication of the state values
- we put explicit arrows towards the second observation variable
- we represent state duration in HSMM of undefined value (dotted line)

P.1.3 *HSMM description (p 6). The definition of a HSMM is too superficial. It should be more concrete/explicit, like for the HMM model. The random variables involved are not described. Also a standard reference like one of these two should be added: Shun-Zheng Yu. Hidden Semi-Markov Models Theory, Algorithms and Applications. Elsevier, 2016 or Barbu, V-S. et N. Limnios. 2008. Semi-Markov Chains and Hidden Semi-Markov Models towards Applications. Springer.*

We have introduced precisions in the presentation of HSMM. In particular, the transition probability in HSMM is now explicit. The PMFs of the sojour time is also more detailed. This indeed required more than small edits in the original text.
It also appeared that the presentation of the geometric PMF was lacking precision. This has been improved and this allowed considering the remark made by reviewer#2 about the definition of the PMFs and about the definition of the Hellinger distance.
We also added the reference to Barbu and Limnios in the reference list.
See Lines 114-142.

P.1.4 *Line 152 : a 'conditionally to the state sequence' is missing*

Correct. This has been specified. Line 153.

P.1.5 *Line 158: do you mean 'by ignoring temporal dependencies in the hidden layer'? There are hidden variables in the mixture model, even if they are independent.*

The hidden variables of a mixture model are the variables defining the various groups. In H(S)MM these groups get temporal coherence bringing some additional information and improving the decoding. The expression "ignoring the temporal dependencies" is probably misleading and has been removed. See lines 161-164.

P.1.6 *Line 218: 'For each case' instead of 'For each model'?*

Yes. Modified.

P.1.7 *Page 10: About the definition of the loss. Since MRA and MSA are probabilities, with values between 0 and 1, wouldn't it be easier to interpret the loss if it was defined as the difference instead of the relative difference?*

We agree with the reviewer that this would lead to a bounded indicator. A potential drawback of the relative difference we used would be when the denominator (the MSA) is close to 0. However this is not the case in the considered scenarios. Therefore, we think that this relative difference indicator remains adapted (and a classical indicator). We want to emphasize that the conclusions of the paper that rely on this indicator remain similar when considering the non-normalized one.

Reviewer 2

Review by Sandra Plancade, 18 Jul 2024 08:40

Overall Comment

This manuscript analyses the impact of model (mis-)specification on decoding accuracy in hidden markov and semi-markov settings, combining supervised experimental data and simulations. The subject is of high interest, and the analyses are conducted rigorously and thoroughly. The choice to restrict yourselves to two settings i.e. two real data sets used to determine realistic parameters for simulations, represents a good compromise to limit the complexity of the results while allowing variations in the setting characteristics. The results are both clear and comprehensive (with a remarkable job to include a large number of characteristics on the same graph while remaining understandable), the main finding is the strong impact of model mispecification on the observed layer on decoding accuracy, versus the weak impact of mispecification in the hidden layer. The discussion represents a true added value to the results. The scope and limits, as well as the summary of the main results, are precise and accurate but the most valuable part is the reflexion on various issues regarding (mis-)modeling, notably based on literature. Even if the present version seems sufficient for a manuscript whose main goal is to provide novel analyses, it would be interesting to provide more details on these aspects.

Major comments

P.2.1 *1) In the discussion section devoted to AR assumptions, you recommend to conduct exploratory analyses of the observation process data to define appropriate auto-correlation hypotheses. If these analyses are straightforward or at least already developped in literature, you should briefly present them and mention alternative models. If not, this issue should be discussed as potential perpectives for further work, if possible evocating leads and difficulties. Moreover, these recommended prior analyses were not conducted on the real data presented in the manuscript. Indeed, the AR0 and AR1 model were chosen a priori, and only an a posteriori analysis on the validity of AR hypotheses is displayed (and concludes to a violation of the AR1 and even the general AR assumption). This appears as a contradiction, and as a minimum, this positionning should be clarified.*

We have addressed this point in the MS and we consider that the points raised here are not fully relevant. As indicated in the MS, generally speaking, the exploratory data analysis cannot be conducted or at least not in a rigorous manner. The key reason is that, given that the states are *unknown*, one cannot explore the pdf of the speeds *by state*, nor the temporal correlations *by state*. We also consider that some exploratory analysis has been done and is being presented in the manuscript. In particular, we provide the correlation coefficients and we mention that none of the pdfs of the speeds are Gaussian while being reasonably symmetrical. If a rigorous exploratory data analysis can not be conducted in general, we however recommend to set hypotheses as much as possible in reference to data analysis. A correlation analysis without knowing the state, does not allow checking the correlation within state *per se*. However, it might provide indications on whether the situation is far or not from an AR0, an AR1, or an auto-regressive process of large order. The remark made by the reviewer tends to indicate that the MS is not precise enough w.r.t this point. We modified accordingly the last sentence of paragraph "Simulation-estimation experiments " of the section "Auto-correlation deteriorates the state decoding accuracy". See lines 467-471.
We also modified the titles of the corresponding sub-sections to make it clearer. Lines 364 + 369.

P.2.2 *2) On real data, accuracy is less good with HSMM-AR1 than expected in a mixture model analysis in setting 2 (Figure 9). This observation is mentioned on l.429-430, but without explanation. The most surprising is that this occures for setting/vessel 2 only, while model mispecification is greater with setting/vessel 1. Do you have an explanation or hypotheses?*

Before answering, we felt the need to clear up a misunderstanding in the reviewer's comment. We believe the reviewer swapped the two boats because the comment is very specific for vessel 1 (settings 1) and not for vessel 2 (settings 2) . On this basis,
i) For real data, accuracy is less good with HSMM-AR1 than expected in a mixture model analysis in

setting **1** (and not in settings 2 as indicated by the reviewer; blue is for setting 1 and red for setting 2).
ii) lines 429-430 concern the model performance under simulations and not the real data.
We rather believe that reviewer 2 drew our attention on lines 461-470 that concern her point.
First, we found that line 460 needed reformulation. The new formulation is now :
For Vessel 1, the most robust model (HSMM - AR1) produces state decoding performances that under-perform the simulation experiments and the mixture model. However, for vessel 2, it produces performances between the simulation experiments and the mixture model. Second, as mentioned by the reviewer, the mispecification is greater with setting 1. Our explanation of the bad performances of HSMM-AR1 for settings 1 are thus provided in lines 468-470 (old numbering). It can be summarized in a confusion in estimated parameters related 1) to less distinguishable speed distributions between the two states and 2) to high variability of the n_s parameter of the Negative Binomial distribution.

P.2.3 *3) I am not sure that the title properly reflects the content of the manuscript. First of all, it seems to me that the terms "decoding", model (mis-)specification, and potentially "(semi-)Markov" should appear in the title. Besides, in the auto-regressive models, the term "long-term correlations" would alude to higher auto-regression levels than the AR1 actually implemented in the manuscript; moreover the AR level is not "properly accounted for" as demonstrated on Figure 14, since the real AR level is much higher than 1. I take the liberty to make a suggestion but in an absolutely non-prescriptive way : "Analysis of model specification in state-space models: account of correlations in the observations improves decoding performances".*

Thanks for these comments and proposal. Indeed the term "long term correlation" is here to alude to higher auto-regression levels than the AR1 actually implemented in the manuscript. We agree that the model implemented are AR0 and AR1 but not more. However, the drastic decrease of decoding performance is likely to be due to long term correlations. We agree that this is not demonstrated strictly speaking but this is what is strongly suggested by our analyses. Accounting though for the suggestion, our counter proposal is
"Proper account of correlations improves decoding performances of state-space (semi) Markov models"

Minor comments

P.2.4 *l.158 (and several other times): It seems that you use the term "mixed model" for "mixture model".*

Good point. Done

P.2.5 *l.160: $\widehat{S}$ could be written explicitely, and the notation s' could be properly defined (even if the latter is intuitive).*

Good point, that indeed concerns also the definition of the accuracy later on in the manuscript. While answering reviewer's point, we also improved the definition of the state estimation and the definition of the accuracy that was not fully accurate in its past form.
Strictly speaking, the states are estimated by maximization of the likelihood. So doing, we do not built an estimator $\hat{S}_t$ function of some *random* variables. We rather search for the value $\hat{s}_t$ that allows maximizing the likelihood; that is the state estimation. In this regards, there is no estimator $\hat{S}_t$ but rather an estimation $\hat{s}_t$. Accuracy is thus now directly defined as the proportion of correct estimations $prop(\hat{s}_t = s_t)$ and not $P(\hat{S} = S)$ which has no sens strictly speaking. See lines 230-231.
The notation s' is now defined. See Line 165.

P.2.6 *l.161: I don't understand what you mean by "empirical speed frequency distributions"; indeed, as far as I undestand, the "true" distribution parameters i.e. the ones used for simulations are actually the empirical ones. Could you clarify this point? Similarly, is there an empirical counterpart for π_s in the formula of d_V?*

Good point. The sentence should not be broken into two parts for better understanding. This has been changed and a full definition for the empirical counter part provided. See lines 168-170.

P.2.7 *l.239: The definition of accuracy is unprecised. If I'm not wrong, S is never defined, does it correspond to a given S_t or to the chain $(S_t)_{t\geq1}$? As the $(\widehat{S}_t)_{t\geq1}$ are not mutually independent,* *formula for the chain $(S_t)_{t\geq 1}$ can not be directly deduced from single S_t formula. Moreover, are the results displayed e.g. in Figure 9 computed with this theoretical accuracy formula or with an empirical counterpart? I recommend to write the exact quantity (either as a sum or an integral) of the computed accuracy.*

We fully agree that it was unprecised (if not more than unprecised). This point is connected with the previous point about the lack of explicit definition for $\hat{S}$. The accuracy has been properly defined as the proportion of correct state estimations. So doing we got one accuracy by simulation-estimation experiment. See line 236.

P.2.8 *l.337: "nearly all parameters are impacted by this bi-layers discrepancy... ". As far as I understand, Figure 12 shows that a non-small proportion of parameters are weakly impacted, notably in setting 1.*

We acknowledge part of the comment here. While the difference between settings 1 and 2 is really not clear, we agree that the sentence is too general wrt to Figure 12. Indeed, parameters related to the first moment of speed distribution are not that impacted. The sentence has been modified accordingly. See line 345.
Caption of Figure 12 has also been improved.

P.2.9 *Figures 14 and 15: provide the definition of the notions (delayed correction plots and coefficient of partial auto-correlation).*

Partial auto-correlation is mentioned in the text. To avoid defining it, we added a reference for readers to search for details if needed. See line 238.
Caption of figure 15 has been expanded to explain what the delayed correlation plots are.

P.2.10 *Considering mixture model as a basis of comparison is very relevant. Compared to the usual distance between PDF (Hellinger, etc), this discrepancy measure has an interpretation in terms of decoding. (This comment does not call for any modification!).*

Thanks. This was exactly the purpose behind this choice.

Typos

P.2.11 *l.165-166: S_{k+1} should be S_{t+1}*
Yes; done

P.2.12 *l.188: I think "by variable (d_{AR_S})" should be "by state (d_{AR_S})"*
Yes; done

P.2.13 *Figure 11 : the x-axis label should be d_{AR} as mentioned on l.322.*
Yes; done. We also improved the figure caption.

Reviewer 3

Summary

This article studies state-space models that arise for instance in ecology to study movement patterns, when these patterns depend on a hidden state that changes over time and that one aims at recovering given tracking data. One possible motivation, which is the focus of this article, is the assessment of fishing pressure on fish stocks using GPS tracking data from fishing vessels, whose movement patterns change depending on whether they are fishing or not. As emphasized by the authors in the discussion, the results of this article are however not restricted at all to this setting. For the applications of interest in this article, a "good" model is one which can recover properly the hidden states given observations, or in other words, a model with a good state decoding accuracy. However, standard methods for model selection are often based on how well the model fits observations rather than the unobserved hidden layer. In this paper, in order to address this issue, the authors use simulation-estimation experiments to assess the state decoding accuracy of four classes of models (HMM versus HSMM, and AR0 versus AR1), and evaluate to what extent they are robust to deviations from model assumptions. These simulation-estimation experiments were performed for two parameter sets inferred from real-world fish vessel GPS tracking data, and the experiments are described in enough details so that they can also be performed for other parameter sets if needed. The main conclusions of the simulation-estimation experiments are that while approximating a HSMM by a HMM does not degrade significantly the state decoding accuracy, the same cannot be said of neglecting auto-correlations. The application to two real-world datasets gives further evidence that these auto-correlations cannot be neglected when trying to estimate hidden states. Evaluation I found the models and methods used very clearly explained, and so despite the fact that they are quite technical by nature. I identified a few inconsistencies in notation that are listed below in the "Minor comments" section, but they do not impede comprehension and should be straightforward to fix. I appreciated that the study was motivated by a clear biological motivation, and that the article alternated between synthetic data and real-world datasets. While some figures are very well-designed (in particular Figure 3) and a very useful addition to the text, others have some design issues (in particular among the ones in the supplementary materials, see the "Minor comments" section below), or could have been put in Appendix instead to reduce the number of figures included in the main article.

P.3.1 *My main major comment on the article is regarding the interpretation of Figures 9 and 10, in the case of Setting 2. A key assumption used throughout the paper is the fact that the HMM and HSMM models considered in this paper are related, in the sense that the HMM is nested in the HSMM. In particular, under this assumption, performing inference under a HSMM when the data follows a HMM should not lead to any decrease of the model accuracy. However, this is not the case in practice in Setting 2, when this leads to a fairly small but noticeable decrease. I agree with the authors that this decrease is small compared to e.g. the effect of neglecting auto-correlations, but it does apparently contradict what seemed to be a key assumption of the study, which guided modelling choices. Since the main topic of this study is to assess the effect of deviations between data a model assumptions 1 on accuracy, I believe this should be commented upon in more details than it is currently the case (L.324-329).*

Although a geomtric PMF is a particular case of a negative binomial PMF, the PMF for the (shifted) negative binomial PMF gets three parameters to estimate when the geometric PMF requires only one parameter. This can be the reason for the slight divergences observed when estimating with an HSMM when the data are HMM. In other words, the HSMM can converge to an HMM but it might not reach it totally. In addition, the n parameter of the negative binomial PMF is not forced to be an integer. It can be a real. So it can tend to 1 without being equal 1. To address this point we have added a new figure (in supplementary material) looking at the n values obtained when inferring with an HSMM while the data are HMM. The objective is to evaluate the capacity of the EM to end up with a negative binomial PMF whose n parameter equals 1. It happens that

- the n of the negative binomial PMF are around 1 with alternate behavior (above/under) given the state
- the shortest-lasting state, i.e. state 1 for settings 1 and state 2 for settings 2, is the one with n smaller than one (and vice versa)
- the fluctuations in the n increases when the data resolution increases, with larger fluctuations for settings 1 where $q_{75\%} = 2$ and n values sometimes equal to 5. The point is that, notably for settings 1, the mean sojourn time in state 2 is very large, with though a very small transition probability.
- there is no difference in the output when using ARO or AR1 structures for the speed variables.

Sentences have been added to (old) lines 324-329 so as a supplementary figure (new figure 15).

Minor comments

P.3.2 *Introduction 1. 1st paragraph You may want to make this paragraph more accessible to a wider audience, for instance by mentioning examples of states of interest that can be recovered from tracking data (like the "fishing/not fishing" states considered in this paper).*

Done.

P.3.3 *2. L.71 "motivated by the fact that a significant proportion"*

Done.

Material and methods

P.3.4 *Figure 1.*

- *"Conditional independence is thus reflected by the absence of an arrow" but there is no arrow between the circles and the triangles The figure has been modified to account for this remark (and a remark from reviewer 1)*
- *Notation not consistent with the one introduced in Table 1. For instance, $V_i(t)$ was denoted $V_{i,t}$ in Table 1, μ_i and σ_i had an additional dependency in the state s_t, and I did not find $\epsilon_i(t)$ introduced in Table 1. That's a good point. It's done and we also added the necessary modifications to Figure 1.*

P.3.5 *L.140 "shift = 1" Shoudn't it be "shift $\geq$ 1" ?*

Done.

P.3.6 *Definition of the Hellinger distance Slight notation inconsistency, earlier "geom($\cdot$)" was used to denote the geometric distribution with probability of success $\cdot$, but here geoms(t) is the probability that a random variable distributed as geom(1 ps,s) is equal to t. Also, the s in sNB might get mixed up with the notation for the state s.*

That's a good point. We clarified the definition of the geometric PMF, and of the shifted negative Binomial as well. We also introduced a dedicated acronym for the shifted negative binomial so that the remark about the 's' notation is solved.

P.3.7 *L.149 "where π_s is the probability of being in state s" when in the invariant distribution for the underlying Markov chain ?*

Yes. We introduced Π for the model and π for the proportion observed over simulations. We modified the text accordingly. See lines 148 and 171.

P.3.8 *P.10, last equation What are R and T ? Moreover, the notation T is already used for the random sojourn times, so I would recommend to choose another notation (this is also valid for Figure 4).*

Notations have been changed in the text and in the figure. See line 239.

Results

P.3.9 *Figure 6 You might want to choose a way to distinguish Settings 1 and 2 that is more visible when printing in black and white (this is a very minor comment though).*

The point is that the figure coding is fixed all along the manuscript. Changing here would require to change everywhere in order to keep this homogeneity. We suggest not to change.

P.3.10 *Figure 9.*

- *What is the meaning of the sign on the upper-left part of the figure ? It is to help reading the figure* from *the simulations in row* to *the estimation in column. If the arrow brings more problem than it resolves, we remove it.*

- *Legend, L.3 ":" at the beginning of the row* Done.
- *Legend, L.5 "the white envelopes" are not white for grey panels* Modified.
- *Legend, L.-3 "the continuous lines" The lines corresponding to the envelopes are also continuous* We added "thick and coloured"
- *Legend, L.-2 two consecutive ":"s* Removed.
- *Legend, L.-1 "sNB" should be "sN B"* Changed.

P.3.11 *L.355 "But choosing a geometric PMF would even be worth it" might be too informal*

This was one of the sentence modified by the professional english proof reader that polish/correct the MS before submission. We believe this correspond to an efficient way of writing the point here.

Discussion

P.3.12 *L.384 "Such a comprehensive range"*

Done.

P.3.13 *L.497-500 This is more a curiosity question than an actual comment, but could it be that the decreasing resolution reduces the autocorrelation and allows one to get closer to the AR0/AR1 case, or not at all ?*

That's a good point. In a sens yes. But the counter part of reducing the resolution is that the number of point by state reduces accordingly, down to a poitn when there is only one point by state or even down to the point that one misses short leaving states.

Appendices

P.3.14 *L.526 "Supplementary figures"*

Done.

P.3.15 *Figure 15 Left part of the legend of the y-axis is cut (part of the S in "State")*

Figure has been adjusted.

P.3.16 *Figure 16*

- *I was unable to read the titles of the y-axes, and I also struggled a lot to read the experiment numbers.* Improvements have been made to the figure.
- *Legend, L.2 The closing parenthesis that should come after "simulation model" is missing.* Done.
- *You might want to skip the mention of the black curve and just state that the EM never reached the maximum possible number of iterations.* Done with max number of iterations indicated

P.3.17 *Figure 17 Same readability issues as for Figure 16*

Done

P.3.18 *L.556 "teeny" might be too informal, you might want to replace it by e.g. "tiny"*

Yes :) , done.