

Review of revised version :

"Proper account of auto-correlations improves decoding performances of state-space (semi) Markov models"

PCI Math Comp Biol

Thank you to the authors for the modifications and the response to the comments. This is a response to the comments not fully addressed by the authors (in orange in their document). I consider that my first major comment has actually been addressed. But regarding the second comment, I am still confused and I recommend additional explanations.

Major comments

1. [My original comment] 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 developed in literature, you should briefly present them and mention alternative models. If not, this issue should be discussed as potential perspectives 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 positioning should be clarified.

[Response of the authors] 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.

[My response to the response] Thank you for the modifications. Indeed, I did not realise that in the discussion, you were considering the situation where a supervised set of data is available (which is actually the case in your application). It is a strong prerequisite that can not always be satisfied in practical applications. In my view, the precision brought in the MS is essential, but then I consider that my concern has been addressed.

2. [My original comment] 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 1.429-430, but without explanation. The most surprising is that this occurs for setting/vessel 2 only, while model misspecification is greater with setting/vessel 1. Do you have an explanation or hypotheses?

[Response of the authors] 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 3 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 misspecification 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.

[My response to the response] All apologies for the messing up in my comment about vessel 1/2 and the line numbers, and thank you very much for having put it back in order.

Nevertheless, I am not fully convinced by your answer. Indeed, my question was about the under-performances of HSMM-AR1 with respect to the mixture model for vessel 1 (blue), in Figure 9, bottom-right. However, your first argument of less distinguishable speed distributions between the two states explains the worse decoding performances on data from vessel 1 w.r.t. vessel 2, **but I don't understand how it is related to the superiority of mixture model on HSMM-AR1 for vessel 1. I consider that this point should be clarified.**

Your second argument mentioned in your answer above, namely the high variability of the n_s parameter of the NB distribution that quantifies the deviation from the geometric distribution, seems different from the one presented in the MS (old numbering 468-470 / new numbering 479-481) related to the AR order. I probably have misunderstood something...

Moreover, I wonder if the arguments suggested in the following lines (from new numbering 482) in the form of a recommendation, namely a misspecification in the correlation between observations that goes beyond AR order, are part of the explanation of the comparative performances of HSMM-AR1 and mixture model.

More generally, my questioning probably originate from a lack of understanding about the role of the comparison with the mixture model. Indeed, while the models HMM/HSMM-AR0/AR1 are partially nested, **I don't understand precisely the positioning of the mixture model. If it is explained somewhere else in the MS, I apologize for having missed it but it would be relevant to refer to it when analysing the performances of the mixture model (around lines 471-...).** If not, I recommend to clarify this point.