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

by N. Bez, P. Gloaguen, M.-P. Etienne, R. Joo, S. Lanco, E. Rivot, E. Walker, M. Woillez and S. Mahévas

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.

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

on accuracy, I believe this should be commented upon in more details than it is currently the case (L.324-329).

Minor comments

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).

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

Material and methods

3. FIGURE 1

   - "Conditional independence is thus reflected by the absence of an arrow" but there is no arrow between the circles and the triangles
   - 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.

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

5. 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 $geom_s(t)$ is the probability that a random variable distributed as geom$(1 - p_{s,s})$ is equal to t. Also, the s in $s\mathcal{NB}$ might get mixed up with the notation for the state s.

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

7. 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).

Results

8. 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).

9. FIGURE 9

   - What is the meaning of the sign on the upper-left part of the figure ?
   - LEGEND, L.3 ":" at the beginning of the row
   - LEGEND, L.5 "the white envelopes" are not white for grey panels
   - LEGEND, L.-3 "the continuous lines" The lines corresponding to the envelopes are also continuous
   - LEGEND, L.-2 two consecutive ":"s
   - LEGEND, L.-1 "$s\mathbb{NB}$" should be "$s\mathcal{NB}$"

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

Discussion

11. L.384 "**S**uch a comprehensive range"
12. 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 ?

Appendices

13. L.526 "Supplementary figure**s**"
14. Figure 15 Left part of the legend of the y-axis is cut (part of the S in "State")
15. Figure 16
    - I was unable to read the titles of the y-axes, and I also struggled a lot to read the experiment numbers.
    - Legend, L.2 The closing parenthesis that should come after "simulation model" is missing.
    - 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.
16. Figure 17 Same readability issues as for Figure 16
17. L.556 "teeny" might be too informal, you might want to replace it by e.g. "tiny"

Questionnaire

Title and abstract

Does the title clearly reflect the content of the article? **No** (only gives a restricted picture of the content of the article)

Does the abstract present the main findings of the study? **Yes**

Introduction

Are the research questions/hypotheses/predictions clearly presented? **Yes**

Does the introduction build on relevant research in the field? **Yes**

Materials and methods

Are the methods and analyses sufficiently detailed to allow replication by other researchers? **Yes**

Are the methods and statistical analyses appropriate and well described? **Yes**

Results

In the case of negative results, is there a statistical power analysis (or an adequate Bayesian analysis or equivalence testing)? **Not applicable**

Are the results described and interpreted correctly? **Yes** (though see second paragraph of the "Evaluation" section)

Discussion

Have the authors appropriately emphasized the strengths and limitations of their study / theory / methods / argument? **Yes**

Are the conclusions adequately supported by the results (without overstating the implications of the findings)? **Yes**