

The general scientific aim of this paper is to model ant interaction data in order better to understand how colonies of ants behave in an laboratory setting. In particular, the main aim is to identify times at which the rate of interactions between ants changes from a low value to a high value, and vice versa. The points at which this appears to happen are evident in plots of the data, so there is adequate information in the data to be able to do this, provided that a reasonable model is used. A statistical aim is to explore the use of penalized inference in the context of a particular model based on a simple stochastic process.

I have considered the content of this paper carefully. It considers an interesting application and there is a substantial amount of work involved in the technical implementation of the statistical methodology that has been used. I'm very concerned about the rationale behind the use of regularization/penalization here. Penalization can be useful to constrain a very flexible model to be desirably smooth, or to input prior information about certain model parameters when this is lacking in the data and otherwise results would be scientifically unrealistic. However, in this paper it is used to try to overcome a fundamental inadequacy of a (hidden Markov) model: it is incapable of reproducing the behaviour that is evident in the data and therefore the results that the authors want. Very strong prior information is input in a rather arbitrary way. Therefore, the statistical aim is inappropriate in this context and the scientific aim is not achieved because, by using a model that is very wrong, we learn little about how ants behave.

It is clear from the outset that the data behave in a way that is very much at odds with the structure of the hidden Markov model described in Section 2. For example, inspection of Figure 1 suggests that the apparent durations in the low intensity state don't behave like a random sample from a geometric distribution. In other words, it's clear that the (first-order) Markov assumption that is assumed to govern the transitions between the states is very inappropriate. Rather than considering how the model could be modified in order to make it more realistic very strong priors are used to force the model to approximate the results that are (clearly) to be desired based on an informal inspection of the data. This involves intensive computations to tackle what at first sight appears like a relatively simple problem.

I don't think that we learn much from this exercise other than that the model is so far wrong that we need to constrain it so strongly to get the results that we want about the transitions between the states. Assessment of the results is based on a judgement of whether the estimated transitions between the states make sense. There is no other assessment of the fit of the model to the data, although in the current setup this would be blurred greatly by the results being dominated by the use of very strong prior information. Interpretation of the results is limited to inspection of posterior summaries of the rates of ant interaction in the estimated low and high intensity states.

There may be some merit in investigating this kind of model but, quite reasonably, the current paper is focussed only on a particular application and offers no more general assessment of the methodology that would provide support for its use in a different setting.

Specific comments

~~page 2, line +4.~~ What is “biologically reasonable” seems to be based on the appearance of Figure 1. If there is some other basis for this then it should be explained.

page 3, line +6. There is no clear explanation, here or in the caption to Figure 1, of the meaning of, and basis for, the red and blue shading. I appreciate the more information is given but even then the rule used to determine red and blue is not explained precisely.

~~page 3, line +13.~~ Are ants allowed to forage in the experiment? The setup of the experiment could be clearer in this regard, particularly since the entrance of foraging ants is involved in the covariate used later in the paper.

page 4, line +12. The 1-step ahead prediction may be better after the penalization (because the model is so wrong) but is it good in absolute terms?

page 5, line -5. It makes some sense to use a Poisson distribution but the vast majority of the counts in the data are in 0, 1. Perhaps we would lose little, and gain some simplicity, in modelling only presence and absence of ant interactions.

~~page 8, line +7.~~ No justification is given for the choice of the prior parameters. In particular, the parameter values in the priors that relate to the components of the transition matrix \mathbf{P} , that is, the vectors (12000, 1) and (1, 12000) result in marginal priors for the diagonal elements (the probability of staying in the same state) that are have Beta(12000, 1) distributions. Therefore, these priors have a mode at 1 and are very strongly peaked at 1. This constitutes very strong prior information, used to try force, in a somewhat arbitrary way, the posterior distribution of \mathbf{P} to place most probability density on matrices with diagonal elements close to 1. How did you decide on the use of 12000, as opposed to some other large number? ... by looking at the results that are produced?

~~page 8, line -11.~~ “confirmed” is a strong claim. I don’t expect that there is a problem but there are some fairly standard and long-standing methods to assess convergence more formally, based on comparing variation with and between more than one chain. I’d be more convinced by these than “visual inspection” alone. Is there any burn-in period?

page 8, line -10. 7 hours is rather a long time. I appreciate that these things can take time but this seems rather prohibitive, especially given that at least some of the analyses presented later in the paper presumably take even longer. How long do the latter take? Why give a timing only in Section 2?

~~page 8, line -7.~~ On what are the colours in Figure 1 based? The MCMC scheme gives many posterior samples of the latent states \mathbf{X} . I expect that you have used the marginal posterior mode at each time point, that is, coloured the time point according to the predominant state in the posterior samples at that time. It would be better to add to the plot the estimated posterior probability of being in the high (or low if you prefer) intensity state. Then we could see whether the model suggests a particular state strongly, e.g. 0.99 vs 0.01, or weakly, e.g. 0.51 vs 0.49.

page 9, equation (15) are these element-by-element posterior means? $\hat{\mathbf{P}}$ is (very close to) being symmetric and $\hat{\boldsymbol{\delta}}$ very close to being (1/2, 1/2). This is not the case in the results in Section 3. Have you thought about why this is?

page 9, page 10. The main point of this section is to impose an even stronger penalty on switching states. Why, in addition, take a slightly different modelling approach to Section 2, i.e. now based on a continuous time Markov chain? In particular, the need to check that the transition matrix in (19) will contain legitimate probabilities is a little inelegant,

even if it works in practice.

page 15, line +9. Viewing the results for the 2-state model in Figure 2 is made more difficult by the inclusion of the results of the 3-state model. This makes it look like choice of τ above e^{-10} give similar results. Is there anything that we can conclude from the MSPE values about how good the predictions are in absolute terms?

page 17, line +10. The switches do align better with our visual impression of the data but we haven't gained much. The results are driven by the prior which prevents meaningful interpretation based on the hidden Markov model, which would say that waiting times in each state are very very flat geometric (or perhaps exponential) distributions, something that isn't consistent with the data.

page 21, line -9. You find no evidence but you don't have anything more than a subjective impression on which to base this. The assessment and interpretation of covariates is potentially an important feature of your approach. Not having a means of assessing the contribution of covariates, for example, via an appropriate measure of goodness-of-fit, is a major drawback.

Data and R code

Thank you for supplying data and code. I have looked at the data in the .csv file and the R code in the .Rmd file. One problem is that (as far as I can see) your bespoke R functions, e.g. `sumvis_troph()` have not been supplied. Also, initially I thought that the numbers of (pairs of) ant interactions in the data don't tally with the information in the Figures that there are fewer than 250 interaction. However, I think it must be that only data from Location=1 are analysed.

Minor comments and typos

The figure captions are too brief. Many of them don't give enough specific information about what appears in the plots.

~~page 4, line +5.~~ Replace "right before" by "immediately before".

~~page 5, line -11.~~ What is N_t ? You haven't defined it explicitly (yet).

~~page 5, line -3.~~ t not t (and elsewhere)

~~page 5, line -1.~~ It would be better if X_t , and it's statespace, were defined earlier, nearer the start of Section 2.1

~~page 6, line +4.~~ How are these gamma random variables related? Are they independent?

~~page 6, line -6.~~ Equation (7) use a comma not a full stop (and "where" not "Where").

~~page 6, lines -4 and -6.~~ Are you missing tildas on the N s?

page 8, lines -7 to -1. What do these posterior distributions look like? Perhaps they could be summarized in a plot?

page 9, line +3. Is the δ notation necessary? I don't think that it is used elsewhere?

~~page 10, line +8.~~ fee ding.

~~page 10, line -6.~~ This equation disappears off the end of the page. Here, and elsewhere, I'm not keen on the use of * to denote multiplication. I would remove these and rely on the convention that multiplication is implied.

~~page 15, line -1.~~ What is \mathbf{N} ? ...and should there a a square in here?

~~page 17, lines -6 and -7.~~ Hats in the wrong places.

~~page 20, equation (50).~~ Is there any rationale behind these priors?

page 21, line +1. Section 3?

~~page 21, "...contained zero".~~ You need to be more careful with your description. Presumably, you mean to refer to the credible intervals.

page 21, lines +6 to +14. You have quoted many numbers but you have not provided any interpretation of them.

page 24. Is the Shaby and Wells (2010) paper still under review? Perhaps cite the technical report.

~~page 27, caption to Figure 3.~~ "enalized".